European Journal of Personality, *Eur. J. Pers.* **29**: 250–271 (2015)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.1986

Lab and/or Field? Measuring Personality Processes and Their Social Consequences

CORNELIA WRZUS[1]* and MATTHIAS R. MEHL[2]

[1]*Johannes Gutenberg University, Mainz, Germany*
[2]*University of Arizona, Tucson, AZ, USA*

Abstract: How can researchers study personality processes and their social consequences? In our methodology overview, we first introduce ambulatory assessment methods, which repeatedly measure experiences, physiology and behaviour in people's daily lives based on real-time assessments of self-reports, physiological activity and behavioural observations. Then, we describe methods suitable for assessing personality processes in laboratory settings: self-reports on interpersonal perception, physiological measurements and behavioural observation. We discuss the combination of field and laboratory assessment methods based on their respective strengths and limitations and then highlight ethical issues surrounding the use of these methods. Finally, we propose future avenues for how developments in mobile technology can be used to advance personality research. The increasing availability and the decreasing costs of smartphones, wearable sensors and Internet connectivity offer unique potentials for further understanding the processes underlying how personality exerts broad and important social consequences. Copyright © 2015 European Association of Personality Psychology

Key words: personality measurement; mobile sensing; experience sampling; ambulatory physiological assessment; behavioural observation

Personality has broad social consequences: For example, more extraverted people are more popular among their peers and have more friends (Asendorpf & Wilpers, 1998; Back, Schmukle & Egloff, 2011c; Ilmarinen et al., 2015; Ozer & Benet-Martinez, 2006); more emotionally stable and agreeable people have more satisfying and durable romantic relationships (Ozer & Benet-Martinez, 2006; Roberts et al., 2007; Schaffhuser et al., 2014). Yet, *how* do more extraverted people attain higher popularity and acquire more friends, and how do more emotionally stable and agreeable people achieve more durable partnerships? The processes underlying personality's social consequences have only recently started to come into scientific focus. Our goal here is to review available methods and provide suggestions on how to capture these processes *in vivo*, which is in people's natural daily environments, and *ex vivo*, which is in controlled laboratory settings.

Personality traits are internal dispositions that manifest in processes: to think, feel, or act in certain ways in specific situations and with intended outcomes (Cervone, 2005; Denissen & Penke, 2008; Fleeson & Jayawickreme, in press). To produce social effects (i.e., influence others), traits have to manifest in behaviour, which in turn has to be perceived by others (Back et al., 2011a; Brunswik, 1956; Funder, 1999). Accordingly, personality processes refer to both behaviour and perceptions (thoughts and feelings), which can be measured using self-report, physiological assessment and behavioural observation.

Throughout the article, we illustrate the described methods using a research example. Specifically, we focus on how extraversion predicts friendship development. Studies often measure trait extraversion at one point in time and predict the concurrent or future popularity and social network size (e.g., number of friends; Asendorpf & Wilpers, 1998, Paunonen, 2003). Here, we provide suggestions on how one could expand on these studies focusing on the underlying processes of how traits (e.g., extraversion) afford social consequences over time (e.g., friendship development). First, studies point to the importance of communication style, positive affect (e.g., laughter and positive facial expressions) and self-confident behaviour to explain the link between extraversion and friendship development (Back et al., 2011c).

In the first section, we introduce ambulatory assessment methods to measure ongoing, everyday experience, physiology and behaviour using (close to) real-time assessment methods. The second section describes self-report, physiological assessment and behavioural observation with a focus on aspects that are specific to laboratory settings. In both sections, we place slightly more emphasis on behavioural observation methods because behaviour is of seminal importance in linking personality to social consequences as personality characteristics have to manifest in observable behaviour to evoke reactions in others. Third, we discuss combining field and laboratory assessments and raise awareness to potential complications. In the fourth section, we highlight ethical issues surrounding the use of specific methods. Finally, we suggest future avenues on how

*Correspondence to: Cornelia Wrzus, Psychological Institute, Johannes Gutenberg University Mainz, 55122 Mainz, Germany.
E-mail: wrzus@uni-mainz.de

Received 29 July 2014
Revised 25 January 2015, Accepted 25 January 2015

technological advancements can be used to advance personality research. Table 1 summarizes the main points and resources of each section. Selected hardware and software solutions are named throughout the text.

FIELD

Assessments in everyday settings have three main advantages (Mehl & Conner, 2012; Trull & Ebner-Priemer, 2014). First, daily life assessments maximize generalizability because they assess participants within the natural pursuit of their lives. Hence, the results should apply to a broad range of contexts (as they were obtained from a range of contexts) and transfer more easily to real-life relative to laboratory findings (Conner, Barrett, Tugade, & Tenner, 2007; Conner & Mehl, in press; Raento, Oulasvirta, & Eagle, 2009; Reis, 2008, 2012). Furthermore, difficult-to-manipulate and/or unethical-to-manipulate phenomena, such as relationship breakups or convalescence, can be studied as they naturally occur in daily life (Sbarra, 2006). Second, most daily life assessments focus on momentary behaviour, thoughts and feelings as opposed to retrospective, assumed or global assessments. Momentary data thus can reduce biases that may diminish the accuracy of self-reports and that arise from incorrect recall, estimation or reports of phenomena (Bradburn et al., 2004; Brose et al., 2013; Schwarz & Oyserman, 2001). Third, repeated daily life assessments uniquely allow studying the prevalence and unfolding of phenomena over time (Bolger et al., 2003; Conner & Mehl, in press; Reis & Gable, 2000). Most phenomena, such as friendship development or marital dissolution, occur over longer time spans that typically cannot be covered in laboratory studies. Thus, ambulatory assessments are well suited to study the psychological processes underlying personality's social consequences. We next consider the sampling of participants and situations (i.e. assessment points) in general, before explaining ambulatory methods for self-report, physiological assessment and behavioural observation.

Sampling of participants—individuals, dyads or groups

Ambulatory assessment studies typically follow a sample of independently recruited participants. However, dyads can also be assessed repeatedly in daily life (e.g. Rauers, Blanke, & Riediger, 2013; Wilhelm & Perrez, 2004), to obtain both the self- and the other perspective on the focal phenomena (e.g. positive affect or the pleasantness of interactions). Complex sample structures (e.g. groups such as people belonging to multiple work units) are possible but require the following: (i) a reliable matching of participants' responses to those of their corresponding interaction partners and (ii) the employment of appropriate data analytic tools (Nestler, Grimm, & Schönbrodt, 2015). For example, friendship development could be studied in several classes of students. Participants' (verbal) behaviour would be assessed in interactions, and participants would rate the interactions and perceived friendship quality with specific interaction partner. The interaction partner needs to be identified within the sample, and future interactions with the same partner need to be tracked and analysed over time. Potential solutions are as follows: (i) presenting participants a predefined, finite (dropdown) list of possible interaction partners (e.g. initials, pictures), from which participants can select the partners they interacted with (Geukes et al., 2015); (ii) using synchronized assessments and/or time stamps of assessments (e.g. Rauers et al., 2013); and (iii) using Bluetooth-transmitted device information on proximate devices (Rachuri & Mascolo, 2011).

Compared with assessing individuals, assessing social groups with three and more interaction partners rating each other offers the advantage of extracting actor, partner and relationship effects in interactions (Kenny, 1994). For example, do more extraverted individuals generally talk more than others (actor effect, Mehl et al., 2006); do specific individuals generally elicit more talking in others (partner effect); do specific dyads differ in their level of talking (relationship effect)? Tracking these effects over time through repeated ambulatory assessments could offer unique insights into how extraversion shapes friendship development through social interaction quantity and positivity.

Sample selectivity and attrition are especially relevant for demanding ambulatory assessment studies. Participants have to repeatedly answer questions in their daily life and/or wear sensors that assess their physiology or behaviour and may restrict aspects of their daily routine. Accordingly, some people may avoid participation (i.e. leading to sample selectivity) or discontinue participation (i.e. leading to selective attrition). Currently, little systematic research on the selectivity and attrition rates of ambulatory assessment studies exists. We know of two exceptions: The sample (age 14–83 years) of a study on situational fluctuations in working memory performance was comparable with a representative panel sample regarding the fluid cognitive capacity (Riediger et al., 2014). Still, samples of exclusively older participants may often have somewhat higher cognitive and physical functioning relative to the general population (Chuiet al., 2014). Selectivity may be reduced by adjusting the following: (i) the recruitment procedure (e.g. emphasizing the study importance, offering compensation and inviting participants from underrepresented groups) and (ii) the number and length of assessments to the target population (e.g. fewer and shorter assessments for time-conscious populations). Attrition may be reduced by strengthening the commitment to the study (personal contact and relevance of the study) and/or through extra compensation for completing all study assessments. Needless to say, researchers should aim for large samples ($n > 200$–300) to have adequate power for detecting both sample selectivity (e.g. in comparison with panel study samples) and individual differences in personality processes, that is, cross-level interactions (Bolger, Stadler, & Laurenceau, 2012; Finkel, Eastwick, & Reis, 2015; Schönbrodt & Perugini, 2013).

Sampling of situations

In addition to sampling of participants, assessments are samples from the larger 'ecologies' of all situations the

DOI: 10.1002/per

Table 1. Overview of main topics and resources for field and laboratory personality assessments

	Implementation	Methodological challenges	Example references	Further Resources
Field				
Self-report	Time-contingent and/or event-contingent assessments once or multiple times per day using smartphones, tablets, and so on	Self-report biases can occur; correlational data offer limited causal inferences	(Bolger & Laurenceau, 2013; Gosling & Johnson, 2010; Hektner et al., 2007; Kubiak & Krog, 2012; Mehl & Conner, 2012)	http://www.ambulatory- assessment.org; http://www.websm.org
Physiological assessment	External or smartphone sensors collect physiological data (e.g. cardiac, electrodermal, physical, muscle or cortical activity)	Validity of smartphone sensors; error sources are difficult to control/assess in daily life	Bussmann, Ebner-Priemer, & Fahrenberg, 2009; Ebner-Priemer & Kubiak, 2007; Myrtek, Foerster, & Brügner, 2001; Wilhelm, Grossman & Müller, 2012	http://www.ambulatory- assessment .org; http://www.openmhealth.org[1]; Bioharness (BIOPAC Inc.), Varioport (Becker Meditec), Vitaport (TEMEC Instruments Inc.)
Behavioural observation	External or smartphone sensors collect environmental or device-usage i nformation (e.g. audio, visual, GPS and user logs)	Participant privacy and data confidentiality; psychologically relevant environmental features need to be established	Chittaranjan, Blom, & Gatica-Perez, 2013; Lane et al., 2014; Lathia et al., 2013; Mehl & Conner, 2012; Mehl et al., 2012; Miller, 2012	http://www.ambulatory-assessment.org; http://quantifiedself.com[2]; http://www.funf.org; http://emotionsense.org
Lab				
Self-report	Perceptions of own and others personality, affective, cognitive, or behavioural states; paper-and-pencil or computer-based ratings	Multiple perceivers (e.g. in round-robin designs) to disentangle perceiver, partner and relationship effects	Bakeman & Gottman, 1997; Biel & Gatica-Perez, 2013; Funder, 1999; Kenny, 1994	
Physiological assessment	Multivariate assessments of neurophysiological, genetic, hormonal, peripheral-physiological assessments, or myography	Appropriate tasks and baselines; comparable assessment conditions, that is, time of day, physical environment and food intake	Cacioppo, Tassinary & Berntson, 2007; Dufner et al., 2014; Kogan et al., 2014; Kreibig, 2010	http://www.psychophys.com/company.html
Behavioural observation	Audiovisual recording of one or more people (individual, dyadic and round-robin designs); standardized coding of people's behaviour	Interaction situations should allow naturalistic behaviour; behaviour coding is resource demanding	Bakeman, 2000; Furr, & Funder, 2007	Mangold Interact; Noldus Observer
Combination of lab and field	Combination of lab and field components within the same study and with the same participants	Lab and field procedures may act as interventions; results may differ, although similar methods are used		
Ethical issues	Informed consent	Privacy protection, handling of sensitive information; data confidentiality	King, 2011; Miller, 2012; Shilton, 2009	
Future directions	Expansion in measures, time and samples	Broad availability of internet access and of smartphones with various sensors that allow long-term assessments together with intensive, (nearly) continuous assessments for several days	Miller, 2012	

[1]The nonprofit start-up Open mHealth is a platform to connect clinical experts with developers of apps and devices.
[2]The *quantified self* initiative is an international movement of people interested in self-observation using technical devices.

participants experience. Researchers can chose between different assessment protocols (Conner et al., 2007; Conner & Lehman, 2012): Ambulatory assessments can occur in response to specific events (i.e. *event-contingent sampling*, e.g. after personal interactions), following a fixed schedule (i.e. *interval-contingent sampling*, e.g. every 3 h), at (pseudo-) random times (e.g. every 3 h+/- a random number of minutes), continuously (mainly for physiological or behavioural assessments) or following combinations of these schedules (Hektner, Schmidt, & Czikszentmihalyi, 2007; Moskowitz & Sadikaj, 2012; Sadikaj et al., 2015; Wheeler & Reis, 1991).

The decision depends on the (assumed) time scale of the focal phenomena: the frequency (e.g. interpersonal conflicts tend to be infrequent) and the continuity (e.g. emotional states are thought to be continuous). Low frequency states (e.g. quarrels) require a higher sampling rate or event-contingent sampling, and longer total duration compared with very frequent states (e.g. positive mood). Complex, contextualized or 'molar' behaviours (e.g. a romantic date) require longer assessment durations (per assessment) to detect than simple, discrete, 'molecular' behaviours (e.g. laughing). Therefore, assessments should be frequent enough to capture change in the focal phenomenon yet short enough to minimize participant burden. For example, physiological assessments can constrain participants in the natural pursuit of their daily lives: Most cardiac or hormonal measurements require temporary abstinence from drinking coffee, smoking or exercising, and placement of body sensors may prohibit taking them off, for example, during exercising, sleeping or showering (Kudielka et al., 2012). Break days with no assessments might help for longer monitoring schedules and minimize overburdening participants. Still, the sampled situations and days should represent the situations, to which researchers want to generalize, for example, weekdays and/or weekends and times when people are typically (not) at work.

So far, no conclusive answer is available on how strongly measurement reactivity occurs for different phenomena (e.g. affective experiences and personality states) depending on the frequency and total duration of ambulatory assessments. First, evidence suggests that with typical momentary assessments on several days, reactivity is often negligible (Barta et al., 2011; Conner & Reid, 2012; see however Larson & Sbarra, 2015).

In addition, event-contingent and time-contingent sampling may suffer from nonrandomly missing data when participants decide not to report specific kinds of situations (e.g. extremely unpleasant or embarrassing situations) or affective states (e.g. intense excitement or anger). We found in our studies that young (adolescent) and less conscientious participants produced slightly more missing data but that the overall amount of missing data was generally small (about 6% of all scheduled assessments). It may be possible to assess situations missing in self-report through observational (e.g. automatic sound recordings) or physiological assessment. Yet, participants always have the right to skip or censor assessments, including the right to turn off automatic behavioural or physiological recordings at desired times.

Self-report

In this section, we describe the assessment of self-reports possible for various psychological phenomena: feelings, thoughts, perceptions of own and others' behaviour and, in addition, the context, in which the psychological phenomena occur in daily life. Self-reports can be assessed in mobile ways using, for example, smartphones, personal digital assistants and tablets, or in stationary ways (also often referred to as online diaries) using traditional desktop computers at participants' work or home. The decision between mobile and/or stationary assessments depends on whether the focal phenomenon occurs repeatedly/continuously throughout the day and is prone to response biases when assessed once per day, for example, in the evening at home (Conner et al., 2007). From an implementation standpoint, mobile and stationary web-based assessments are converging (Bolger & Laurenceau, 2013; Hofmann & Patel, 2014), and we focus on mobile assessments and then briefly cover stationary online diary assessment tools.

Practical issues

In addition to decisions regarding the participants (e.g. individuals or groups of connected people) and the sampling of situations (Section on Sampling Situations), the researcher has to design the questionnaire. The questionnaire should be as short as possible (<5 min) to motivate immediate response. We suggest always assessing contextual information (e.g. main activity, presence of other people and location) because contexts likely vary between participants and within participants over time and can alter the focal phenomenon. For example, the positive affect in conversations with future friends may vary as a function of whether or not other people are present and as a function of whether the setting is private or public. The contextual information are still based on participants' self-reports and therefore reflect their subjective perceptions (e.g. perceived privacy of place) rather than an 'objective' or observer's reality (e.g. actual privacy such as in a closed room). Some lines of research suggest that the subjective, psychological features of situations are more important than physical features and have developed and validated instruments to assess the psychological features of situations (e.g. DIAMOND, Rauthmann et al., 2014; RSQ, Sherman et al., 2013; Wagerman & Funder, 2009; Morse et al., 2015). Often, having both subjective and objective contextual information available can prove valuable. The Section on Behavioural Observation reviews nonself-report-based methods of assessing aspects of participants' momentary environments.

Hardware and software

Decisions on mobile devices and software are crucial and depend on prior decisions regarding study design including assessment schedule, topics/types of data and the longer-term focus of the research agenda. The most important purchase criteria right now are operating platform, screen size for presenting items and stimuli, battery life, connectivity (e.g., Wi-Fi, Bluetooth) and built-in or connectable sensors such as GPS, audio and video recording (Section on Behavioural

DOI: 10.1002/per

Observation). For a single study, renting devices together with preinstalled software may be reasonable [currently possible with, e.g. mQuest (http://www.mquest.eu), SIS (http://www.sismarketresearch.com) or movisensXS (http://xs.movisens.com)]. Another elegant, flexible and low-cost solution is to use participants' own smartphones but may bear (at least) two risks: The appearance and perhaps functionality may differ between devices—with possibly detrimental effects for psychological studies, which seek maximally comparable study conditions. Samples may be selective, if characteristics of smartphone owners differ by social stratum, (sub-) culture or even personality, yet, the increasing prevalence of smartphones worldwide (http://www.portioresearch.com) should gradually deemphasize these methodological concerns.

Typically, participants receive the (random) triggers via email or text messaging on their smartphone and follow a link to an online questionnaire (e.g. http://www.surveysignal.com, Hofmann & Patel, 2014; http://www.soscisurvey.de is free for research purposes). In addition, some open source software (e.g. ohmage, Hicks et al., 2011; Ramanathan et al., 2012), commercial apps (which may be used for research purposes) and commercial software (movisensXS, http://xs.movisens.com) remind participants directly to report events and experiences. Conner (2014); Conner and Mehl (in press), Miller (2012) and Kubiak and Krog (2012) provide overviews of currently available free and commercial software solutions. These overviews detail the operating platform, available item and output formats, possible assessment schedules, online data processing capability and further special features for each software. Also, the Society of Ambulatory Assessment (http://www.ambulatory-assessment.org) provides regularly updated information on technical solutions.

Stationary field assessment

Stationary, internet-based assessments at participants' home, or online diaries, have become highly convenient over the last few years. As in ambulatory assessments, the sampling protocol, questionnaire and technical issues need to be decided (Gunthert & Wenze, 2012). Naturally, home-based, stationary online diary assessments can only occur once or twice per day, but they can be implemented over a longer time span than typical momentary experience sampling assessments, and it is possible to have more in-depth assessments (i.e. administer more items or questionnaires)—but participant burden must still be considered. There is no general rule on maximum duration per day or in total because what is acceptable depends on the nature of the sample. Often email reminders are used to maximize compliance, yet, participants and researchers should exert precautions to bypass spam filters (e.g. by including the sender in the address book). Several commercial and open-source services are available to set up an online diary study quickly and to collect data efficiently (e.g. Amazon's mTurk: Boynton & Richman, 2014; Buhrmester, Kwang, & Gosling, 2011; SosciSurvey: Leiner, 2014; see also Hewson & Lason, 2008 or http://www.websm.org for a broad range of software solutions). Also, researchers may decide to program their study independently (Fraley, 2004; Gosling & Johnson, 2010) with consequently full control of layout, items formats, functionality and data storage as not all available software includes all special features (e.g. reaction time experiments, automatic feedback and presentation or assessment of audiovisual content).

Summary

In sum, self-report assessments in daily life have become relatively easy to implement with a variety of technical solutions to choose from. They allow access to phenomena that are otherwise difficult to measure. However, ultimately, they reflect the subjective perception of feelings, thoughts and behaviour. Next, we present methods to assess physiological parameters that are relevant for studying personality processes.

Physiological assessment

A broad range of physiological measures can be assessed in daily life: hormonal activity from saliva samples (Hofman, 2001; Schlotz, 2012), cardiac activity (e.g. heart rate, blood pressure and stroke volume), respiration patterns, electrodermal activity, body temperature, muscle activity (e.g. facial muscles), physical activity, eye movements (Ebner-Priemer & Kubiak, 2007; Wilhelm, Grossman, & Müller, 2012) and even cortical activity (e.g. Kranczioch et al., 2014) can be measured repeatedly or continuously in daily life. For example, measures of cardiac or electrodermal activity could serve as indicators of emotional experiences (Kreibig, 2010) during social interactions, which differ between zero acquaintances and long-term friends. Measures of eye movement or physical activity of the head could serve as indicators of attentional focus during conversations.

Practical issues

Similar to self-report assessments, researchers have to make important decisions about the assessment protocol, that is, the sampling schedule and the total assessment duration. Continuous assessment of physiological activity seems the logical default option but has two drawbacks. First, very large amounts of raw data accumulate per participant even over short sampling periods (e.g. about 150 MB for 24 h-ambulatory ECG and accelerometry). A possible solution could be to discard the raw sensor readings and to store only preprocessed data, if available for the focal measure (e.g. from ECG or accelerometry, automatically computed heart rate or activity counts can be stored and require less space, Bussmann et al., 2009; Ebner-Priemer & Kubiak, 2007). A disadvantage of this approach is that outliers cannot be classified as valid or erroneous recordings, and often, outliers are discarded already during preprocessing. Second, most physiological data can be unambiguously interpreted only when sufficient further information (e.g. on the context) are available. For example, peaks or changes in heart rate can only be related to emotional experiences if information on other factors that influence heart rate are obtained, such as physical activity, body posture, time of day, environmental noise,

DOI: 10.1002/per

medication and stimulants (Schlotz, 2012; Wilhelm et al., 2012).

Possible solutions could be to implement time-contingent or event-contingent sampling (combined with context-sensitive sampling, Intille, 2012) for physiological data and assess necessary contextual and experiential information for the same period. For example, in a study on emotional experiences during interactions with friends, researchers could decide to record cardiovascular activity only when microphones detect that participants are speaking with friends (or others, respectively, as control condition). If information from other sources (e.g. smartphones) and physiological data should be linked, researchers must verify that the devices are synchronized before recording. Importantly, most devices tend to desynchronize over longer periods, which varies across devices, unfortunately.

Hardware and software

In addition to less expensive single-channel systems (e.g. for blood pressure Accutracker II, Suntech Medical Instr.; for accelerometry Actigraph, ActiGraph LLC), multichannel devices (e.g. Bioharness, BIOPAC Inc.; Varioport, Becker Meditec Inc.; Vitaport, TEMEC Instruments Inc.; VU-AMS, http://www.vu-ams.nl) offer a broader range of functions, such as concurrent recording of multiple physiological parameters, sound and environmental conditions. For detailed comparisons of specific devices, see recent overviews (e.g. http://www.ambulatory-assessment.org, de Vries et al., 2006; Ebner-Priemer & Kubiak, 2007; Wilhelm et al., 2012).

Furthermore, smartphones possess more and more built-in sensors (e.g. accelerometry, barometer, temperature and photoplethysmography via camera sensors, Miller, 2012; Scully et al., 2012) and are increasingly available to complement or substitute stand-alone solutions. Unfortunately, information on the reliability and precision of smartphone sensors and their measurements are often missing (Rachuri & Mascolo, 2011; Raento et al., 2009). Researchers should establish validity in pilot studies by comparing smartphone-derived measures against established ambulatory devices (e.g. see before).

Most commercial devices come with specialized software (e.g. Variograph for Varioport, Vitascore for Vitaport) for data collection that also facilitate data preprocessing (e.g. detrending and outlier detection). Often some data analyses are also possible (e.g. computation of distributional characteristics and spectral analyses).

Summary

In sum, physiological assessments can complement self-reports by supplying additional indicators for the focal latent phenomenon (e.g. emotion), which may capture the dynamic changes over time more precisely than self-reports, and are less susceptible to reporting biases. Yet, physiological assessments in daily life often contain variance from other sources than the focal phenomenon because the experimental control—available in the laboratory through limiting location, physical activity and food and beverage intake—is missing. Behavioural observation data seem to suffer somewhat less from this problem of unwarranted variance. Furthermore, behavioural observation is of critical importance for studying the social effects of personality at the process levels because personality characteristics have to manifest in observable behaviour to elicit consequences in other people. Next, we discuss ambulatory methods for behavioural observation in daily life.

Behavioural observation

Ambulatory assessment methods that allow for the direct, observational assessment of behaviours (e.g. speech and movement) in naturalistic settings in daily life are a recent development relative to experience sampling and physiological ambulatory monitoring. At the same time, the future likely holds the biggest scientific leaps forward in this domain given the surge in technical progress in smartphone and wearable sensing technologies (Chittaranjan, Blom, & Gatica-Perez, 2013; Lathia et al., 2013; Miller, 2012; Raento et al., 2009).

Developed at the end of the last century, the Electronically Activated Recorder (EAR) intermittently records snippets of ambient sounds and thereby creates an acoustic log of a participant's day as it unfolds (Mehl et al., 2001; Mehl, Robbins, & Deters, 2012). By now, the EAR is an established method for the unobtrusive observation of real-world social behaviour. Participants carry an iEAR device on them as they go about their normal lives. Past research has consistently yielded valid assessments with 'thin slices' of audio (e.g. 30 or 50 s), a small number of recordings per hour (e.g. one every 9 or 12 min) and a total monitoring period of 2–4 days (Mehl & Holleran, 2007).

In a second step, the ambient sound recordings are then coded for aspects of participants' moment-to-moment locations (e.g. in a public or private place), activities (e.g. watching TV or eating), interactions (e.g. alone, in a group or on the phone) and emotional expressions (e.g. laughing or sighing). So far, EAR research has relied entirely on manual labelling of the sound files by trained coders. Although it is already feasible to automatically detect a limited number of sound-based behaviours (Lane et al., 2014; Lu, Pan, Lane, Choudhury & Campbell, 2009; Rahman et al., 2014), the validity of such computerized codings in different contexts and for different populations still needs to be thoroughly tested. Also, personality studies so far have had a tendency to be interested in broad sets of potential behavioural manifestations (Mairesse et al., 2007; Mehl et al., 2006) rather than a narrow set of target behaviours for which computer scientists are conducting first validations studies (e.g. social engagement; Harari et al., 2014).

Prior research with the EAR method supports the idea that real-world observational data can yield findings that may otherwise be difficult or impossible to obtain. For example, in a cross-cultural study, Ramirez-Esparza and colleagues compared self-reported sociability to sociability observed with the EAR in the USA and Mexico (Ramírez-Esparza et al., 2009). They found that although American participants *rated* themselves significantly *more talkative* than Mexicans, they actually spent almost 10% *less* time talking. In a similar way, Mehl and colleagues found no

DOI: 10.1002/per

gender difference in EAR-recorded daily word use despite strong stereotypes that women are more talkative than men and despite the fact that women report being more talkative than men in personality questionnaires (Mehl et al., 2007).

Researchers have further developed other ways of assessing behaviour directly and unobtrusively in the real world. For example, participants' location can be tracked via Wi-Fi and GPS information (Montoliu et al., 2013; Wolf & Jacobs, 2010). Depending on the density of Wi-Fi hotspots and cell towers, the location of mobile phones can be estimated with the precision of several metres or yards, with the limitation that GPS usually works worse within buildings. Having a detailed log of GPS coordinates is often of little psychological use. What researchers typically need is meaningfully labelled location information (e.g. at home or at work), which so far requires at least partial manual user labelling (Do & Gatica-Perez, 2013). Such location information then enables researches and practitioners to better understand and perhaps even alter everyday behaviour, for instance, preventing alcohol or drug abuse in high-risk locations such as around bars (Epstein et al., 2014; Gustafson et al., 2014).

Cameras also provide objective information on participants' location and context. Devices like the Narrative Clip, a wearable lifelogging camera (http://getnarrative.com) or Google Glass (http://www.google.com/glass/) can take pictures or videos of participants' surroundings and sort (and maybe even annotate) them automatically. Research using these new, currently underdevelopment technologies just started (e.g. Wettstein & Jakob, 2010; see also Rosalind Picard at the MIT) and carries with it some delicate ethical issues (Section on Ethical Issues). However, at the same time, it also holds considerable promise for furthering researchers' understanding of social situations (Rauthmann et al., 2014).

Clearly, the future will hold important developments in the area of mobile sensing, that is, the 'reading' of behaviour patterns, emotions and environments from people's phone usage (Chittaranjan et al., 2013; Eagle & Pentland, 2005; Onnela, Waber, Pentland, Schnorf, & Lazer, 2014). For example, de Montjoye, Quoidbach, Robic and Pentland (2013) showed that the personality of smartphone users (e.g. extraversion or neuroticism) can be derived from the log files of the phone usage (e.g. number of interactions, number and diversity of contacts, response latency to events and distance travelled). The Section on Hardware and Software provides further examples on mobile sensing.

It is easy to envision how mobile sensing bears the potential for a major leap forward in the study of personality processes. Using our example of extraversion and friendship development, a researcher might, for example, recruit groups of loose acquaintances (e.g. university freshmen in a dorm) into a longitudinal study where their task would consist of hardly more than living their normal lives. The degree and nature of their daily interactions could be extracted from the unobtrusively monitored audio stream surrounding them, Bluetooth proximity (in-person conversation) or calling logs. The content of text messages (Underwood et al., 2012) and social media postings (Qiu et al., 2012) could be analysed from phone logs. Other contextual aspects could be extracted from labelled location information (e.g. at a friend's place) and classification of the ambient sounds (e.g. one-on-one or group setting). The development of friendship quality could be assessed directly via experience sampling or indirectly via extraction of paraverbal information (e.g. tone of voice), body posture and movement (e.g. via accelerometry) and facial expressions (e.g. extracted from wearable camera pictures). Finally, a social network analysis of all interactions could reveal the user's popularity. Of course, much of this currently falls under the rubric science fiction, and some aspects are disconcerting and highlight the need for the field to develop strong privacy and data confidentiality guidelines (King, 2011; de Montjoye, Wang, Pentland, Anh, & Datta, 2012).

Practical issues

Practical issue in the context of behavioural observation largely revolves around technology issues. First and foremost, battery life is a crucial point for intensive long-term assessments (Rachuri, Efstratiou, Leontiadis, Mascolo, & Rentfrow, 2013). Both stand-alone devices and smartphones may last not much longer than a day if applications run continuously. For example, when Bluetooth is used to detect whether two devices of new friends are close, the continuous Bluetooth scanning quickly drains the battery. Continuous GPS tracking recording also uses much energy. Context-sensitive sampling, piggybacking on other programs that already run in the background, and machine-learned smart sampling (Miller, 2012; Rachuri & Mascolo, 2011) could solve some of the energy issues. At the moment, participants should be instructed and ideally automatically reminded (e.g. in movisense software) to recharge the device overnight.

Another practical issue concerns the rapid development and high turnover in smartphone and wearable technologies. As methodologically desirable as it is, it can often be impossible for researchers to complete a project with the same generation of a mobile device that they started out with. With large ambulatory assessment studies often running for several years but with new versions of mobile devices and operating systems being released on a yearly basis, researchers can find themselves forced to update devices and/or operating systems in the middle of the study. Such transitions can affect the reliability of the employed sensors and software and can thereby lead to critical disruptions in the data collection. Such transitions further often change peripheral yet nontrivial aspects of the data collection such as the size and weight of the mobile devices that participants carry around (e.g. researchers tend to seek the smallest possible device to minimize participant burden, whereas consumers often favour large screens) and the format in which the data are collected (e.g. file format, recording quality and sensor precision). A potential solution is to deliberately 'overpurchase' devices at study onset to ensure that enough will survive attrition (wear-and-tear or theft) until the end.

Hardware and software

Conner and Mehl (in press) provide an overview of selected behavioural observation systems. In addition, the website of

DOI: 10.1002/per

the *Society for Ambulatory Assessment* (http://www.ambulatory-assessment.org) maintains resource pages for hardware and software (app) solutions. Most apps run on iOS and/or Android devices and smartphones that fulfil the technical requirements (e.g. regarding sensors). For example, the current EAR system, the *iEAR*, consists of a free iOS app that runs on iPod touch and iPhone devices. The visual observation systems *Narrative Clip* and *Google Glass* can be ordered directly from the company websites (respectively, http://www.getnarrative.com and http://www.google.com/glass).

Apps often can be downloaded directly from the AppStore or GooglePlay store. For example, *Funf* (http://www.funf.org) allows to collect information on location, communication behaviour, physical activity, ambient sounds and general usage of the mobile phone. The *StressSense* app (Lu et al., 2012) monitors ambient sounds for the user's voice, extracts stress-relevant parameters (e.g. speech rate, pitch variability and jitter) and integrates the parameters into stress estimates based on machine-learning algorithms (calibrated against users' skin conductance). In a similar way, Rachuri et al. (2013) developed *Emotion Sense*, a mobile phone application for the automatic recognition of discrete emotions based on voice parameters. Finally, Lane et al. (2014) developed *BeWell*, a smartphone application that monitors and provides feedback on users' physical activity (via the embedded accelerometer), sleep activity (via the accelerometer and recharging information) and social activity (via ambient sound monitoring). The *BeWell* app is currently being adapted for and used in personality research (as part of the *Dartmouth Biorhythm Project*; Harari et al., 2014).

Summary

Despite the recent progress in this area, behaviour assessment in naturalistic, everyday settings continues to be the Achilles' heel in personality field research. Whereas experience sampling is now cheap, reliable and easy to implement for everyone and physiological ambulatory monitoring has become affordable and feasible for many, naturalistic behaviour observation remains effectively off limits for most researchers. It either requires having the resources to deal with very large amounts of behavioural coding (e.g. using the EAR method) or it necessitates building collaborations with computer scientists to implement state-of-the-art sensing methods (Frauendorfer et al., 2014). In addition, even the most advanced social sensing tools still struggle with extracting information at a level that is useful for personality researchers. For example, knowing how many active Bluetooth connections surround a user at a given time provides rather rudimentary information given that many people surrounding the participant may not have (activated) Bluetooth connectivity. Furthermore, Bluetooth connections are silent about the psychological ties they electronically represent (e.g. stranger, friend, coworker or partner). Yet, because of the important role that observed behaviour plays in the process of personality expression and perception, it is clear that the future of personality process analysis lies, at least in part, in smartphones and wearable sensors and that, over the years to come, important developments can be expected to come out of this assessment domain.

LAB

Laboratory assessments have two main advantages for the study of personality processes and their social consequences compared with daily life assessments. First, the range and quality of accessible data are higher in the laboratory compared with daily life. For example, it is possible to obtain direct and more detailed audiovisual recordings of people's behaviour. In contrast, ambulatory assessments are somewhat restricted regarding the temporal coverage and content, typically covering only short snippets and small ranges of behavioural observation (e.g. vocal recordings and physical proximity between interaction partners).

Second, laboratory assessments usually offer more control over the following: (i) the context; (ii) possible disturbances; and (iii) potentially the behaviour of interaction partners, which might also affect the social outcomes. For example, when studying effects of low extraversion on social behaviour in conversations to predict friendship development, naturalistic observation can be difficult for the following reasons: (i) if some participants never enter contexts to meet new people during the study period (e.g. parties); (ii) if a third person disrupts a conversation; or (iii) if potential friends undermine conversations. Researchers also have more control regarding the selection of all participants and their assortment, whereas interaction partners of participants usually cannot be controlled in ambulatory assessments. We next address the sampling of participants and situations in laboratory settings (i.e. the study design).

Sampling of participants—individual, dyads or groups

In addition to observing individuals or dyads (e.g. one friendship pair), researchers can employ different round-robin designs (Back & Kenny, 2010; Kenny, 1988). For example, in a round-robin design with indistinguishable dyads, everyone interacts with everyone else (e.g. zero acquaintances assigned to the same discussion group). In round-robin designs with distinguishable dyads, everybody interacts with everybody else, but the dyads differ qualitatively (e.g. mother-oldest child, father-oldest child, etc. Kashy & Kenny, 1990). In full-block designs, members of one group interact with all members of another group (e.g. in speed-dating paradigms, all women interact with all men; Asendorpf et al., 2011; Berrios, Todderdell, & Niven, 2015; Finkel, Eastwick, & Matthews, 2007). Importantly, participants in round-robin designs can interact in groups (e.g. in group discussions) or dyads (e.g. in private conversations). As a critical advantage, round-robin and full-block designs can distinguish actor (target), partner (perceiver) and relationship effects, which are inseparable when studying dyads (Back et al., 2011a; Kenny, 1994). Compared with ambulatory assessments of groups, laboratory round-robin and full-block designs allow for the creation of groups of comparable size and controlling the duration of the interaction. In ambulatory assessments, some participants might interact with none or only one of the other participants rendering social relations analyses impossible for these participants (Kenny, 1994).

DOI: 10.1002/per

Sampling of situations

Even if behaviour is elicited in the laboratory, the situations can be designed to resemble real-world contexts in psychologically important ways (referred to as *experimental realism*). For example, the testing room can be set up to resemble a living room (for observing parent–child or spousal interactions, Goodwin, 2012, or a café/bar, Stappenbeck & Fromme, 2014). Unobtrusively placed cameras and microphones further enhance the naturalistic impression of the context (Goodwin, 2012) and reduce participants' reactance to observational equipment (Brockner, 1979). Yet, efforts to enhance the naturalism of the context can interfere with data quality, such as when microphones are too far away to record low voices or when cameras cannot capture the participants with sufficient precision.

The tasks can also resemble everyday activities and obviously have to reliably elicit behaviour relevant for the focal research question (cf. *relevance* Funder, 1995; Funder, 1999). Researchers can adopt—and if necessary adapt—existing paradigms such as couple or family interaction tasks (Vater & Schröder-Abé, 2015; Weiss & Heyman, 1990), speed-dating protocols (Asendorpf et al., 2011; Berrios et al., 2015; Finkel & Eastwick, 2008; Finkel et al., 2007), self-introductions to peers (Back et al., 2011c), dyadic (Aron et al., 1997; Funder & Sneed, 1993; Ickes, 1983; 1993) or triadic (Letzring et al., 2006) interaction tasks, group discussions (Hall & Watson, 1970; John & Robins, 1994; Kenny, 1994, 2004) or other problem-solving (e.g., Robins & Beer, 2001) or stress tasks (Back et al., 2009; Borkenau et al., 2004; Kirschbaum et al., 1993). Otherwise, researchers should carefully pilot the task with members of the target population to detect potential pitfalls. In general, investigators need to decide on whether typical (i.e. average) or extreme (i.e. maximal) behaviour is of interest and whether behaviour should be elicited naturally or experimentally manipulated. And, careful theoretical consideration should be given to select situations that reliably evoke individual differences in trait-relevant behaviour. For example, for studying the expression and perception of neuroticism, it is important to employ tasks that are moderately stressful (Hirschmüller et al., in press). Similarly, for studying processes invoking openness to experiences, experimental situations should afford the expression of personal interests and preferences (Rentfrow & Gosling, 2006; Stopfer et al., 2014).

For example, for studying effects of extraversion and conversation behaviour on friendship development, the researcher could elicit conversations without specifying a topic (Jacobs et al., 2001) or manipulate the emotional valence of conversations through asking participants to converse about either pleasant or conflictual topics (e.g. Gonzaga et al., 2007; Vater & Schröder-Abé, 2015). Experimentally manipulating the positivity/negativity of conversations through the behaviour of confederates (or other influencing factors) can facilitate causal inferences but may create a strong psychological situation that can constrain the expression of individual differences (Cooper & Withey, 2009). Importantly, laboratory assessments typically examine short-term processes in the range of minutes. Hence, repeated assessments might be necessary to examine how personality processes and the social consequences unfold over longer periods.

Self-report

In this section, we discuss the assessment of personality process data collected in the lab using self-report. Important personality processes include personality self-perception, the interpersonal expression and perception of personality and cognitive-affective personality dynamics. Because of space constraints, our review focuses primarily on the study of interpersonal perceptions as one highly active research domain (Funder, 1999; Kenny, 1994).

Interpersonal perceptions in dyadic or group interactions usually focus on general perceptions of the following: (i) the person (e.g. evaluative judgments and specific personality characteristics); (ii) the relationship with the person (e.g. attraction and disliking); or (iii) meta-perceptions (e.g. perceived disliking, such as 'how much might the other person dislike me' and perceived other perception, such as 'how extraverted do I seem to the other person'?). Hence, interpersonal perceptions reflect the subjective reality of people regarding other people compared with behavioural observation that aims for objective information. Sometimes, perceptions of specific behaviour are also examined and become similar to behavioural observation coding described in the Section on Behavioural Observation. In addition, self-perceptions of personality characteristics, affective, cognitive or behavioural states can be assessed equally well.

Practical issues

After the researchers selected the constructs and instruments, participants have to be familiarized with the content and the items of the instruments. For example, in research on the Big Five dimensions, it can happen that researchers and lay people conceptualize openness differently with the latter ones often interpreting openness as being open to people—a method artefact that can undermine the validity of openness judgments.

Ratings can be obtained paper-and-pencil or electronically (via desktop computers, tablets, etc.) one to multiple times during the interaction, immediately after or with some delay to the interaction. Again, researchers need to ensure that ratings are uniquely identified. For example, when judging spontaneous liking after interactions with different individuals, participants need to have unique identifiers, which can be entered on the paper rating sheets or the digital devices (e.g. participants can wear IDs on name tags). Still, researchers must ensure that all ratings are anonymous and will not be shared with the targets, for example, in studies involving romantic partners, colleagues or supervisors and their employees.

An important practical issue in the study of interpersonal perceptions lies in the complexity of round-robin designs where every participant interacts with every other participants. If interpersonal perceptions among unacquainted people are studied, participants have to be kept separated until the experimental tasks start. Waiting in the same room or

DOI: 10.1002/per

viewing later interaction partners ‘across the room’ while interacting with another person (e.g. as during real speed-dating events in cafés) would compromise the intention of studying the first moments of interaction between truly unacquainted people.

Apart from such requirements of the physical set up, an important consideration is that the number of questions to be asked per interaction partner (target) tends to be constrained by the number of targets that are being studied. Comprehensive ratings of person characteristics, relationship qualities or meta-perceptions (e.g. more than a few minutes per target) may not be feasible with 5, 10 or 20 interaction partners. Furthermore, sequence effects regarding targets or questions can arise and should be minimized (e.g. by presenting questions in random order; by randomizing the rating position as in speed-dating paradigms).

Hardware and software

As mentioned before, self-reports can be obtained with no or minimal technical equipment, for example, using paper-and-pencil questionnaires or spreadsheet-based (electronic) scoring sheets, where columns represent questions and rows represent targets. Any commercial computer or mobile device (e.g. tablet) is appropriate, and constraints may arise only from specific, customized software. For example, if changes in self-reported liking during interactions are to be combined with behavioural observation data, the system for recording behaviour (see Section on Behavioural Observation of the Lab part) must be synchronizable with the system used for collecting the interpersonal perceptions.

An innovative method for studying the process of personality perception at large-scale comes out of computer science. Biel and Gatica-Perez (2013) recently explored crowdsourcing personality perceptions via Amazon’s Mechanical Turk. Specifically, they studied more than 400 YouTube vlogs (video logs) or other conversational video blogs, in which users posted personal video testimonies on YouTube. The authors first uploaded short, 1-min vlog slices onto MTurk (an online study platform with several hundred thousands registered participants) and obtained personality annotations by more than 2000 human judges. This corresponds to a novel laboratory setting, where participants do not actually come to the lab, but the study materials and design are nonetheless (somewhat) under the researcher’s control. Importantly, the crowdsourced personality impressions were psychometrically on par (e.g. regarding distributions, intercorrelations and interrater agreement of Big Five traits) with what lab-based studies usually obtain, speaking to MTurk’s potential as a research platform (see Buhrmester et al., 2011; Paolacci & Chandler, 2014, who also discuss limitations of MTurk). In addition, personality perceptions could be traced back to (largely) automatically extracted verbal (e.g. speaking time and prosodic information) and nonverbal (e.g., gaze, pose and facial expressions) cues from the audio and video channels.

Summary

Most researchers are familiar with the assessment of self-reports and many established instruments and guidelines exist (Schwarz & Oyserman, 2001). We therefore focused on interpersonal perceptions as mediating process how personality differences afford social consequences. Importantly, interpersonal perceptions are inherently linked to personality manifestations in observable behaviour (see Lab section on Behavioural Observation). Furthermore, moderators exist for both the manifestation and perception of personality (Funder, 1995; Funder, 1999; Gosling et al., 1998; Letzring, 2008; Letzring, Wells, & Funder, 2006; Vazire, 2010). For example, certain traits such as extraversion manifest more often and more accurately in overt behaviour and hence can facilitate accurate personality perceptions (Gosling et al., 1998; Vazire, 2010). Also, well-adjusted individuals are easier to ‘read’ and judge reliably than less well-adjusted individuals (i.e. less extraverted, agreeable, conscientious and emotionally stabile people). This shows through higher levels of self–peer, peer–peer and peer–behaviour agreement of judgments (Colvin, 1993). Therefore, interpersonal perceptions vary not only with characteristics of the perceiver but also of the perceived one (i.e. target) and their relationship. This calls for round-robin designs capable of distinguishing such effects.

Physiological assessment

Laboratory physiological assessments can broadly be grouped into neurophysiological (e.g. EEG, structural magnetic resonance imaging, functional magnetic resonance imaging and near-infrared spectroscopy), genetic, hormonal (e.g. from blood, saliva and urine), peripheral-physiological (e.g. ECG and electrodermal assessment) assessments and myography (e.g. facial electromyography and electrooculography). All these methods are suitable for studying physiological correlates of personality traits (at least moderately stable individual differences) and/or associated processes during social interactions (fluctuating states). As a comprehensive treatment is not possible here (but see Robins et al., 2009; Schinka, & Velicer, 2012), we focus on methods that have been successfully used in personality research and can be applied within paradigms of naturalistic person-to-person interaction. That means that similar methods as in daily life can be applied (see Field section Physiological Assessment) because people should be able to move and communicate naturalistically. We focus on lab-specific information for hormonal and peripheral-physiological assessments, yet some of the information also applies to myography or neurophysiological assessments.

Practical issues

In contrast to ambulatory assessments that mainly focus on situation related changes, laboratory physiological assessment often focuses on individual differences in the level of physiological functioning. For example, higher average vagal activity has been associated with greater self-control (Geisler & Kubiak, 2009) and with greater prosociality (but too high levels predicted again lower prosociality, Kogan et al., 2014). Similarly, experiences of state happiness have been repeatedly linked to faster breathing, faster heart beat and lower heart rate variability, which also decreased in

DOI: 10.1002/per

studies on experiences of anxiety or anger (Kreibig, 2010). This demonstrates that single physiological indicators are neither unique for traits (or states) nor are specific traits (or states) linked to only one/a few physiological indicators.

In addition to individual differences in average physiological parameters, *changes* in response to controlled tasks (and individual differences therein) are also relevant for personality processes. For example, how is extraversion related to individual differences in physiological reactions to positive or negative social information? Naturally, the chosen laboratory tasks should elicit the focal physiological response (for task overviews, see Diamond & Otter-Henderson, 2007, Westermann et al., 1996). Furthermore, researchers need to assess baseline measures of their physiological variables in addition to measures during the completion of tasks to assess reactivity and to control for individual differences in baseline or average measures. Sometimes, so-called *vanilla* baselines are appropriate as they reflect physiological activity during nondemanding, neutral tasks in the same posture as later used during the experimental task (Jennings et al., 1992). Compared with resting baselines, vanilla baselines tend to be comparable with experimental tasks regarding cognitive demand and physical activity. In addition, they tend to produce more stable within-session and retest measurements (Jennings et al., 1992).

Often the initial reactivity of a physiological system is focused (i.e. mean-level differences between baseline and task measurements), but individual differences in the time-dependent change can be equally interesting indicators (Brosschot & Thayer, 1998; Davidson, 1998; Wrzus et al., 2014). Because most physiological measures can be readily assessed continuously, the course of the physiological response and also of recovery processes can be assessed. Such time-based measures can be partly independent from level-based reactivity measures (e.g. Wrzus et al., 2014).

Compared with self-report assessments, the data collections involving physiological assessments usually take longer because unexpected technical difficulties may occur and group sessions are scarcely possible. Participants have to be thoroughly informed and carefully prepared for physiological measurement (e.g. sensors have to be placed individually by trained experimenters). Thus, the duration of single assessment sessions and the total period of data collection for sufficiently large samples often will be longer compared with less demanding assessment methods. This time needs to be taken into account during project planning and recruitment of participants.

We explained in the beginning of this section that the strength of laboratory assessments lies in the control about the assessment situation and thus potential error sources. Therefore, as many error sources as possible should be controlled and laboratory assessments conditions should be kept highly similar for all participants. For example, assessment should occur at comparable times during the day, in the same room (under the same environmental conditions) and after the same instructions regarding food and other consumable substances because time-of-day, environmental temperature and noise and food, coffee, cigarettes or medication influence hormonal and cardiovascular data (Bonnemeier et al., 2003; Hofman, 2001; Lovallo & Thomas, 2007; Schultheiss, & Stanton, 2009).

Hardware and software

Devices appropriate for ambulatory assessments can also be used in laboratory settings (see Field section on Physiological Assessment). Such devices may be especially well suited because they are relatively unobtrusive and hence facilitate naturalistic social interactions in laboratory paradigms (e.g. group discussions, spousal arguments and cooperative tasks). In addition, comprehensive lab solutions are offered by companies such as BIOPAC Inc. (http://www.biopac.com), Electrical Geodesics Inc. (http://www.egi.com) or ADInstruments (http://www.adinstruments.com) and often include devices for neurophysiological assessments and myography (including eye tracking). Some comprehensive behavioural observation laboratory systems (e.g. from Mangold or Noldus, see Lab section on Behavioural Observation) are also capable of integrating biophysiological data (for more companies, see http://www.psychophys.com/company.html).

Depending on the type of hormone, hormonal samples can be taken invasively from blood samples or noninvasively from saliva, sometimes even urine and hair, but the interpretation differs (e.g. momentary vs. average hormone level over last weeks, Schlotz, 2012). Saliva samples allow easy and reliable assessments of unbound momentary hormone levels that reflect both trait (regular person-specific hormone output) as well as state components (hormonal output due to recent situational demands). Cortisol and testosterone are robust to storage at room temperature for several days, but samples are best kept in a refrigerator, which becomes the only necessary technical device if external professional laboratories are hired for hormone extraction.

As described in the Field section for Physiological Assessment, most technical devices provide their specialized software (e.g. AcqKnowledge for BIOPAC; LabChart for ADInstruments). In addition, data preprocessing and analysis can be performed with multipurpose software such as Matlab, R or most other statistical software.

Summary

The possibilities to assess physiological correlates of personality traits and processes are numerous in laboratory settings. Still, researchers should have assumptions about the physiological systems that are activated during the psychological phenomenon under focus. For example, when studying individual differences in stress reactions during conflicts with friends, researchers might want to assess the function of the hypothalamic–pituitary–adrenal (HPA) axis (Foley & Kirschbaum, 2010) and/or the cardiovascular system (Burg & Pickering, 2011, for overviews on different systems Cacioppo et al., 2007). Because psychological representations (traits or states) are often linked to several physiological indicators, which in turn correlate with several psychological phenomena, multivariate assessments should be the method of choice. Put differently, the different physiological systems (endocrinological, cardiovascular or neurological system) are linked closely in regulatory circles (Berntson & Cacioppo, 2007) and multivariate time series data enable understanding

DOI: 10.1002/per

the complex regulatory links between physiological and psychological manifestations to form models on individual differences in personality processes.

Behavioural observation

Structured behavioural observation in controlled experimental situations has greatly advanced the understanding how personality traits manifested in observable behaviour (e.g. Borkenau & Liebler, 1995; Funder & Sneed, 1993; Kurzius & Borkenau, 2015; Morse, Sauerberger, Todd & Funder, 2015). Observed behaviour is often quantified by trained observes (see Section on Practical Issues for information on observer training) through codings and/or ratings. Codings typically refer to objectively quantifying occurrences of certain behavioural acts, whereas ratings refer to subjectively toned judgments how intensely, and so on, certain behaviour occurred or how the behaviour appeared. In reality, codings possess some subjectivity because some leeway exists whether the occurrence of the behaviour was perceived (i.e. was the smiling behaviour apparent enough as to be perceived by an observer or an automatic cue extraction program, Biel & Gatica-Perez, 2013). Similarly, ratings should be as objective as possible, which is achieved through highly standardized observer manuals with comprehensive examples and through intensive observer training.

We already explained different study designs and tasks, and now, we describe practical issues regarding the data recording, and the behavioural coding and offer a selective overview on currently available hard- and software solutions.

Practical issues

After choosing the design and pretesting the tasks participants have to complete, the observational equipment needs to be purchased and installed (for suggestions see Section on Hardware and Software). With the first behavioural observation data recorded, a coding system should be chosen or developed. Some behavioural cues can be automatically extracted (e.g. features of facial expressions related to emotions, e.g. Biel et al., 2012; Terzis, Moridis, & Economides, 2010; for a commerical solution, see http://www.emotient.com, or verbal information such as speaking time or prosodic information, Biel & Gatica-Perez, 2013; Narayanan & Panayiotis, 2013). Most of the time, experts (i.e. trained researchers) code behavioural information. Currently established coding systems for interpersonal behaviour include the Riverside Behavioural Q-sort (Funder, Furr, & Colvin, 2000), systems to code marital interactions (e.g. Bakeman & Gottman, 1997; Bakeman & Quera, 1995; Weiss & Summers, 1983) or family interactions (Eyberg et al., 2009; Gordis & Margolin, 2001; Robinson & Eyberg, 1981), systems to code facial expressions (e.g. FACS, Ekman & Friesen, 1978). Although it is recommended to decide on or develop the coding system on the basis of theoretical considerations prior to the data collection, in practice, coding systems often need to be adjusted (e.g. categories added, divided or refined) as the actual study data becomes available.

In general, researchers have to decide on four domains with respect to the behavioural coding system (Bakeman, 2000; Furr & Funder, 2009) First, are the codings meant to capture narrow or broad behavioural domains? For example, are only behaviours related to expression of positive affect coded (e.g. smiling, frowning and laughing) or also other aspects of the interaction? Because behaviours are captured on audio or video records, researchers can always go back to the raw recordings and expand the scope of the initial codings. Second, how fine-grained should the codings be? For example, should instances of smiling (molecular) or the general level of positivity (molar) be coded? Are words, sentences or longer statements the appropriate unit for codings? Although molecular codings can be aggregated to represent broader behavioural characteristics, molar codings likely capture the overall behaviour better than the sum of its constituting molecular parts—a phenomenon known as emergence. Often, it can make sense to combine the two and, for example, use general ratings of closeness and specific coding of closeness indicators such as physical contact. Third, what is the (intensity) threshold for coding/interpreting a behaviour? Often, less leeway exists for counting behavioural acts compared with ratings of behaviour, although the decision whether a certain behaviour occurred or not is also partly subjective. Precise coding manuals with several examples reduce the ambiguity and can lead to greater coding accuracy and hence intercoder agreement. For example, if the coding manual specifies in detail semantic or syntactic features of positivity in statements during conversations and provides examples, less interpretation on behalf of the coder is necessary regarding whether and to what extent positivity during conversations occurred. Fourth, researchers need to decide whether to code events, intervals (e.g. every 5 s or 1 min) or sequences of timed events (Bakeman, 2000). In addition to frequency codings, the duration and intensity of behaviours can also be rated. For example, occurrences of positive statements can be coded (frequency-based event codes; e.g. as operationalization of the level of overall positivity). Each minute of a certain interaction can be coded regarding whether or not a positive statement occurred, how long it lasted and how intensely positive it was (interval codes; e.g. as operationalization of the change of positivity over time). Beginnings, ends and intensity of positive statements in two (or more) interaction partners can be coded sequentially (sequential, timed event codes; e.g. as operationalization of dyadic, interdependent development of positivity).

Once the coding manual is established or adjusted, coders (observers) need to be trained [some researchers (Furr & Funder, 2009) suggest selecting coders according to intellectual abilities, personality characteristics (Letzring, 2008), and intrinsic motivation, who will likely provide highly reliable codings]. Coder training usually follows a two-step procedure: During the practice phase, coders first read and understand the coding manual with one selected example that was chosen to be especially instructive regarding unambiguous behaviour and potential pitfalls. Coders then practice coding and rating with a few examples and are encouraged to ask and discuss open questions. During the coder agreement phase, observers code and rate several behavioural recordings from different participants independently. Based on the independent codings and ratings, the coder agreement is

DOI: 10.1002/per

computed (Bakeman, 2000; Furr & Funder, 2009). The coder agreement usually is the main indicator of coding quality because often objective reference criteria (i.e. 'correct solutions') are unavailable. Importantly, the intercoder agreement should be computed regularly throughout the entire coding process to detect drops in coding quality. Also, the agreement should be computed on the level of later analysis, that is, not necessarily on the level of single smiles if an aggregate measure of friendliness is used for associations with personality traits or social consequences.

Finally, behavioural codings can then be statistically related to other information, such as personality reports, interpersonal perception, physiological data and other behaviour codings. For example, more extraverted people show friendlier facial expressions and more creative stories in social introductions to unacquainted people (Back et al., 2011c). Is this behaviour sustained as people become better acquainted? Is positive expressivity related to more or less activation in the peripheral nervous system? Are positive facial expressions answered by reciprocal smiling and increased positive evaluation, as it can be observed in heterosexual (speed-)dating situations (Back et al., 2011b)?

Hardware and software

Depending on the scope of the research project (ranging from a single study to long-term research agendas), researchers' equipment for the behavioural assessment can range from single, relatively simple cameras to complete and customized commercial laboratory solutions (e.g. from Mangold, http://www.mangold-international.com; Noldus, http://www.noldus.com). Clearly, single cameras are cost-efficient and allow behavioural recordings from stationary viewpoints. Data are directly saved on the devices and later transferred to computers for data storage and data analyses. Complete lab solutions can include multiple cameras and microphones that record participants from different perspectives and are directly linked to computers. The computers allow to steer cameras (e.g. zoom-out or change the angle, if participants moved out of the visual field), to store the data and to edit and analyse the data. We generally advise that the number of cameras and microphones and their placement (depending on the visual field and sensitivity) should be carefully considered and tailored to study specifics: Will participants stay within an assigned space or will they move around? Are all important body parts (e.g. face, hands or whole body) captured in sufficient detail? Can the different cameras (as well as the physiological equipment) be easily and accurately time synchronized after the fact or is it necessary to implement external time (e.g. sound) signals? What data formats are available and are they compatible with the intended software for further processing (i.e. editing and coding) of the recordings?

Complete lab solutions usually include the software needed for editing and analysing the behavioural data. *Interact* (Mangold) is a powerful software capable of editing and coding audiovisual data of various file formats. It provides all the options discussed before (and more) regarding implementations of complex behavioural codings systems (e.g. event-based coding, time-based coding and parallel coding of multiple behavioural streams). In addition, specific components visualize biophysiological data simultaneously with behavioural data, and the *p.a.t.t.e.r.n.* component detects behavioural patterns. *The Observer* (Noldus) is equally powerful (Vogel et al. 2012) and additionally offers *Pocket Observer* to code behaviour in real time from observations outside the laboratory. Like *Interact*, it can incorporate physiological data. *Theme* (Noldus) is specialized on detecting larger patterns in behavioural observations, for example, detecting the pattern 'exchange of toys' from a behavioural sequence similar to 'forward movement Child A—placing hand on ball—closing fingers around ball—picking up ball—moving arm towards Child B—Child B moving arm towards hand of Child A', and so on (Magnusson, 2000). *FaceReader* (Noldus) is another specialized software for detecting emotion-related facial expressions in dynamic video recording based on the FACS (Ekman & Friesen, 1978). Similarily, *Emotient* (originally developed as Computer Expression Recognition Toolbox) is a commercial system for automatic FACS coding (http://www.emotient.com). Validation studies have shown high agreement with trained human coders (Terzis et al., 2010; Littlewort et al., 2011) and electromyographical recordings of facial muscles (D'Arcey, 2013). The Generalized Sequential Querier program (Bakeman & Quera, 1995) is a free, Windows-based software that allows conducting sequential behavioural analyses of observational data coded with the sequential data interchange standard (SDIS; Bakeman, 2000; Bakeman & Quera, 1992; Observer and Interact data can be converted into SDIS).

Summary

'Compared to questionnaires, behavioral observation is difficult, expensive, and can also be limited in scope. Therefore, the first step in behavioral assessment should be a careful consideration of the cost and benefits of the method in the relevant research context' (Furr & Funder, 2009, p. 274). We outlined both the costs (time-intensive and resource-intensive and not every personality characteristic can be readily observed under laboratory conditions) and benefits (measures avoid self-report biases and represent actual behaviour). Researchers must weigh these anew for every research question. For domains where self-reports have known limitations and for researchers interested in observable manifestations and social consequences of personality, it is likely a fruitful method.

FIELD AND/OR LAB?

It is a scientific truism that all methods have strengths and weaknesses. Ambulatory assessments often provide causally ambiguous correlational data and sources of 'error' are difficult to control. In addition, ambulatory assessments often measure behaviours of interest in indirect ways (e.g. interpersonal contact inferred from the proximity of two active Bluetooth connections) or in rather distal manifestations (e.g. behavioural residues observed from living rooms

DOI: 10.1002/per

photographs). Laboratory studies have a better handle of these issues, often though at the expense of limiting the generalizability of findings because the focal phenomenon is studied over a short time scale and within a contrived context. It seems logical to develop study designs that integrate both ambulatory and experimental laboratory assessments.

An 'aiming-for-ideal' study for our example might measure extraversion among unacquainted individuals and observe social interactions in small groups of four to six people to allow variance decomposition of behaviour into actor, partner and relationship-specific variance. In addition to cardiovascular and electrodermal assessments (related to affective experiences), behavioural codings and ratings from trained coders could cover, for example, the amount of interaction between interaction partners (e.g. speech duration and eye contact), the positivity of the behaviour (e.g. smiling and laughter) and the assertiveness (loudness of voice, word use and body posture). Social consequences (e.g. liking and intentions for future joint leisure activities) could be measured before, during and after the interaction using self-reports.

After releasing participants from the lab, the small groups of potential friends could be followed up in daily life using context-sensitive ambulatory assessments. Ambulatory assessments could capture cardiovascular and physical activity, audiovisual recording and self-reported liking and friendship intensity every time the smartphones detects another smartphone user from the same original small group through Bluetooth—and perhaps as control condition, times when participants from other small groups are present. Behavioural observation could be coded regarding speech duration, word use, smiling, laughter and eye contact to maximize the connection to the laboratory observations. Further laboratory assessments, interventions and personality trait assessments for detecting personality change and reciprocal influences among personality and friendships (Neyer & Asendorpf, 2001; Mund & Neyer, 2014) could complement the study. Such a design would allow examining how individual differences in extraversion manifest in emotional and behavioural processes during social interactions and how such processes influence friendship development and potentially change extraversion in the long run (and a range of mediator and moderator analyses). It would help to better understand, which characteristics of social interactions (number, positivity and activity) best predict friendship development or loneliness, respectively.

Clearly, such a complex study would need very large financial and personnel resources to achieve a large enough sample with sufficient power (Bolger et al., 2012). Often behavioural observation or experience sampling studies include relatively small samples. Yet, recent studies show that sufficiently large and heterogenous samples are possible when using participants' own smartphone (Mappiness project, http://www.mappiness.org.uk, MacKerron & Mourato, 2013; Social Fabric project, http://socialfabric.ku.dk). We advocate the use of such integrated designs yet also point out two potential challenges.

First, ambulatory and lab-based assessments may not always correspond strongly. For example, when assessing differences between interactions with friends and strangers in daily life and during lab-elicited conversations, low correspondence may emerge despite employing similar (e.g. coding) methods. The correlation between daily life and lab-based assessments may be low for several reasons, among them that some people may differentiate strongly between friends and strangers in their daily lives but follow the salient norm or expectations (i.e. demand characteristics) for the prescribed interaction in the laboratory. Also, because the laboratory assessment often occurs only once, situational factors, such as being unusually quiet or talkative on a certain day or even time of day (Fleeson, 2001; Hasler et al., 2008), affect laboratory assessments more strongly than ambulatory assessments, where situational factors are assumed to vary and level out over the assessment period. As these influences vary between persons, they can dampen the convergence between ambulatory and lab-based results. Nonetheless, of course, discrepancies are often theoretically informative in that they point to potential moderators of the studied phenomenon.

Second, both repeated ambulatory assessments and intensive laboratory assessments run the risk of unintentionally acting as intervention, which can change subsequent assessments (Larson & Sbarra, 2015). For example, repeated reports and even unobtrusive observations of social interactions in daily life might induce reflections about one's relationships, yet, so may an intensive laboratory session. Costly solutions could be the following: (i) to include control groups that participate only in baseline and follow-up measurements, but not in the ambulatory and laboratory assessments, or (ii) to counterbalance the order of ambulatory and laboratory assessments. This may however not always be possible because of resource (even larger samples are needed) or practical constraints, for example, when laboratory assessments provide baseline measures for friendship quality, which later predict behaviour and changes in friendship quality in daily life.

In sum, the decision for ambulatory and/or laboratory assessments will depend on the research question. For studying the processes of personality's social consequences, repeated assessments are highly desirable. Technological progress has made the implementation of ambulatory assessments easier and more cost-effective. At the same time, repeated assessments over long periods in controlled yet realistic laboratory contexts (e.g. labs that resemble homes or cafés) can be possible alternatives or complements. In both cases, the rich, longitudinal, multivariate (and multitarget) data provide much needed information on the processes underlying how personality influences everyday thoughts, feelings and behaviour and thereby has social consequences, that is, influences people's social relationships.

ETHICAL ISSUES

Collecting real-life data from multiple sources and in several situations prompts (at least) four ethical and legal risks: information about the participants that they did not want to reveal is obtained, information about others is unintentionally

DOI: 10.1002/per

obtained, information about medical illnesses or criminal acts is obtained and any of the before mentioned information becomes available to third persons. Here, our goal is primarily to bring awareness to these issues rather than to provide fully satisfying solutions.

With respect to participants, researchers can limit the amount and type of data to what is essential for the current research question. Also, participants should receive sufficient information, on the nature of the collected data and its potential (mis)use, to provide truly informed consent. This is especially difficult when using innovative technologies, such as Google Glass, where data protection risks are unclear.

With respect to others, unawareness of data collection is problematic. For example, as of now, some US American states (Pennsylvania) and countries (e.g. Germany) do not allow the use of the EAR method because of the possibility of recording identifiable conversation partners and bystanders without their informed consent. However, some places with two-party consent laws, such as California, have deemed EAR research legal and ethical to the extent that people around the participant are made aware of the possibility of being recorded (e.g. via visual and verbal recording alerts). An additional solution to protect the privacy of participants, conversation partners and bystanders is transforming voices (e.g. pitch transformation with the iEAR app) permanently and prior to being stored on the mobile device. This maximizes third-party confidentiality while preserving important paraverbal and linguistic cues for decoding gender, emotion and personality.

Yet, even approved data collection might prompt considerations for action. For example, ECG recordings might detect cardiac arrhythmia for the first time for a person, or behavioural observation with voice recorders or cameras might record criminal acts. Regarding medical information, researchers must consider the reliability of the recordings especially in daily life and can consult clinicians before talking to the participant. Because participants also have a right to 'not know', researchers should ask the participants during obtaining consent, whether participants would like to learn about abnormalities. Regarding criminal information, researcher must check the local laws whether they are obliged to report cases of potential criminal acts to the respective authorities. In the USA, researchers can apply for a Certificate of Confidentiality with the National Institutes of Health, which protects identifiable research information from forced disclosure in civil, criminal or administrative matters at the federal, state or local level. The protection afforded by a Certificate of Confidentiality, though, does not affect the investigators' legal responsibility to report, based on state laws, known or suspected child and elderly abuse and an imminent danger of harm to self or others. Again, whatever the specific risks and protections in a given study are, participants need to be made aware of them prior to the study onset.

Novel ethical challenges are encountered and novel solutions to the protection of human subjects need to be developed in the context of *big data* that are collected in joint ventures between academic researchers and researchers at for-profit companies. Universities and private companies tend to have different motives regarding the use of the collected data and they adhere to different regulations (e.g., nonfederally funded research conducted by private companies is not subject to regulation by institutional review boards). Therefore, delicate ethical situations can arise when academic researchers work with data sets that are collected by IT giants like Google, Microsoft, Facebook, Twitter and telecommunication companies like AT&T, Verizon or Telecom. One such collaborative enterprise between Facebook and academic researchers recently stirred up a massive online ethical debate (Kramer et al., 2014) around the requirement for informed consent and the use of personal information for research purposes. On the positive side, this case brought much necessary attention to the following fact: (i) better oversight over non-academic research is needed and (ii) that existing standards for human subject protection were not designed for these scientifically attractive half-academic-half-business big data research scenarios, and often, the standards have limited applicability (Watts, 9 July 2014). As a related issue, data security must be guaranteed when transferring (e.g. using SSL encryption for data up and download) and saving participants' 'anonymized' data accessible to third parties (e.g. on commercial servers/clouds). Previous work showed that in anonymized datasets participants are individually identifiable with minimal information (Sweeney, 1997).

In sum, researchers can take several measures to protect the confidentiality of their data and the anonymity of their participants. However, with many different data sources (e.g. GPS and time allows inferences about absence from work) or data that is by nature not deidentifiable (e.g. video recordings), specific measures need to be taken. Possible solutions could be to extract relevant information directly on the device (e.g. for speech Lane et al., 2014; Onnela et al., 2014), conduct only local computation on the mobile device (e.g. computing parameters regarding covariance of focal phenomena, such as content of conversation and liking of conversation partner) and discard all raw data immediately (King, 2011; Miller, 2012; Shilton, 2009). Personality researchers, including ourselves, love rich data sets that can provide a more comprehensive picture of participants. Such broad information is often critical for better understanding the focal phenomenon. However, because our research targets are human lives, special measures need to be taken to ensure the physical, social and psychological well-being of the participants.

FUTURE DIRECTIONS—EXPANSION IN MEASURES, TIME AND SAMPLES

What will the future of studying personality's social consequences hold? It is probably safe to say that the future of personality research will see a substantial broadening of measures. Given the immense progress in smartphone technology and sensor, this broadening is bound to first and foremost affect the *in vivo* assessment of behavioural and physiological parameters. Already now, it is possible to extract selected activity and speech-related behaviours automatically from people's normal smartphone usage

DOI: 10.1002/per

(de Montjoye et al., 2013; Lane et al., 2014). Clearly, the precision of the sensors and extraction algorithms will increase over time (e.g. automatic speech recognition and feature extraction; Stark et al., 2014; asssement of cardiovascular activity from blood-flow induced micro changes in facial skin colour, Hernandez et al., 2014). With respect to behavioural observation, algorithms might rival the validity of equivalent human codings at some point. In addition, with the onset of the era of wearable technologies, extraction of a much broader range of behaviours will become feasible (e.g. health-relevant nonspeech body sounds such as eating, drinking or coughing; Rahman et al., 2014).

From the perspective of personality psychology, we imagine that the biggest leap forward may come from advancements in the measurement of environments. It is the measurement of situations that historically has been the sore spot of personality (and social) psychology (Funder, 2006; Rauthmann et al., 2014), and it is there where novel measurement approaches may have the biggest leverage. With the array of environmental sensors that are on board of any new smartphones (microphones, cameras and Bluetooth) and through combining mobile sensor (e.g. GPS) with geographic information system data (e.g. temperature, weather and neighbourhood characteristics), it should be possible to characterize people's moment-to-moment physical and social environments with unprecedented depth and resolution. This, in turn, will allow to develop, test and refine transactional models of personality that are sensitive to different types of person × situation interactions (Neyer & Asendorpf, 2001; Neyer et al., 2013; Rauthmann et al., 2015) and where each component of the *personality triad*, which is personality, situations and behaviour, can be assessed with independent methods (Funder, 2006).

In all of this technology excitement, though, it is important to not relegate traditional experience sampling to 'second class data'. It is tempting to think that the quantified personality processes ought to be only objectively, digitally measured and that subjective reports are second best to sensor reports. Yet, from a theoretical perspective, it will always remain necessary to distinguish the actor's from the observer's perspective, and the subjective perception from the objective reality. Therefore, we are convinced that classical experience sampling is not only here to stay but should also ideally be a standard component of any mobile sensing study (Wang et al., 2014).

Beyond the broadening of measures, though, we think that an important future direction for the study of personality's social consequences is to try to integrate three different time scales that so far have largely been studied separately (and, more generally, to incorporate time at a theoretical level; Luhmann et al., 2014). Because of practical and methodological constraints, most personality studies have so far either focused on studying how personality effects unfold over minutes to hours (in the lab), over days to weeks (experience sampling and daily diaries) or over months to years (longitudinal studies). Clearly, though, personality effects happen at all three levels—and they happen at all three levels nonindependently. Measurement burst designs can help integrate distal processes that unfold at slower time scales and proximal processes that unfold at faster time scales (Nesselroade, 1991). One way to accomplish such a time scale merger would be to create opportunities to tag lab-based and ambulatory assessments onto existing longitudinal studies. Another way would be to extend personality studies that originally were designed as one-time (lab or ambulatory) assessments and convert them into actual longitudinal studies. Both approaches should be highly attractive from a scientific perspective but both are also highly resource intensive and therefore require appropriate and dedicated funding mechanisms. Finally, of course, novel statistical tools need to be developed to synthesize personality processes at these different time scales (Ram et al., 2014, Nestler et al., 2015).

Last but not least, personality psychology would also do well to broaden its samples with respect to participants' sociodemographic characteristics (Henrich et al., 2010). The decreasing costs and the increasing distribution of smartphones and wearable or built-in sensors allow to study heterogeneous samples, for example, regarding age, ethnic or economic background or health status. Including diverse samples or clinical samples with physical or mental illnesses can become easier when more and more people own smartphones and can instal the study software on their phone (e.g. http://www.mappiness.org.uk, MacKerron & Mourato, 2013). Such online-only studies with smartphones could even reach participants in countries with fewer computers and researchers and would enhance representative cross-cultural research. Diverse samples would allow to investigate which personality effects are unique to certain (age) groups (Wrzus et al., 2015) and how social or cultural contexts magnify or constrain personality effects (Gebauer et al., 2012; Jokela et al., 2015; Reitz et al., 2014). Thus, more heterogeneous and representative samples provide a stronger basis for generalizing the findings of personality research.

ACKNOWLEDGEMENTS

The authors thank Annette Brose and Christine Gutsch for their valuable comments on drafts of this paper. The preparation of this manuscript was supported by NIH grants 5R01AT004698 and 1R21HD078778 to Matthias Mehl.

REFERENCES

Aron, A., Melinat, E., Aron, E. N., Vallone, R. D., & Bator, R. J. (1997). The experimental generation of interpersonal closeness: A procedure and some preliminary findings. *Personality and Social Psychology Bulletin*, *23*, 363–377.

Asendorpf, J. B., Penke, L. & Back, M. D. (2011). From dating to mating and relating: Predictors of initial and long-term outcomes of speed-dating in a community sample. *European Journal of Personality*, *25*, 16–30.

Asendorpf, J., & Wilpers, S. (1998). Personality effects on social relationships. *Journal of Personality and Social Psychology*, *74*, 1531–1544.

Back, M. D., Baumert, A., Denissen, J. J. A., Hartung, F.-M., Penke, L., Schmukle, S. C.,…Wrzus, C. (2011). PERSOC: A unified framework for understanding the dynamic interplay of

DOI: 10.1002/per

personality and social relationships. *European Journal of Personality*, *25*, 90–107. DOI: 10.1002/per.811

Back, M. D., & Kenny, D. A. (2010). The social relations model: How to understand dyadic processes. *Social and Personality Psychology Compass*, *4*, 855–870.

Back, M. D., Penke, L., Schmukle, S. C., Sachse, K., Borkenau, P., & Asendorpf, J. (2011). Why mate choices are not as reciprocal as we assume: The role of personality, flirting, and physical attractiveness. *European Journal of Personality*, *25*, 120–132. DOI: 10.1002/per.806

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*, 533–548. DOI: 10.1037/a0016229

Back, M. D., Schmukle, S. C., & Egloff, B. (2011). A closer look at first sight: Social relations lens model analysis of personality and interpersonal attraction at zero acquaintance. *European Journal of Personality*, *25*, 25–238. DOI: 10.1002/per.790

Bakeman, R. (2000). Behavioral observation and coding. In H. T. Reis & C. M. Judd (Eds.), *Handbook of research methods in social and personality psychology*. (pp. 138–159). New York: Cambridge University Press.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. New York: Cambridge University Press.

Bakeman, R., & Quera, V. (1992). SDIS: A sequential data interchange standard. *Behavioral Research Methods*, *24*, 554–559.

Bakeman, R., & Quera, V. (1995). *Analyzing interaction: Sequential analysis with SDIS and GSEQ*. New York: Cambridge University Press.

Barta, W. D., Tennen, H., & Litt, M. D. (2011). Measurement reactivity in diary research. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 108–123). New York, NY: Guilford Press.

Berrios, R., Totterdell, P., & Niven, K. (2015). Why do you make us feel good? Correlates and interpersonal consequences of affective presence in speed-dating. *European Journal of Personality*, *29*, 72–82. DOI: 10.1002/per.1944

Berntson, G. G. & Cacioppo, J. T. (2007). Integrative physiology: Homeostasis, allostasis, and the orchestration of systemic physiology. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.). *Handbook of psychophysiology* (2nd ed., 433–452). Cambridge, England: Cambridge University Press.

Biel, J., & Gatica-Perez, D. (2013). The YouTube lens: Crowdsourced personality impressions and audiovisual analysis of vlogs. *Multimedia, IEEE Transactions on*, *15*, 41–55.

Biel, J. Teijero-Mosquera, L. & Gatica-Perez, D.. (2012). *FaceTube: Predicting personality from facial expressions of emotion in online conversational video*. ICMI'12, October 22–26, 2012, Santa Monica, California, USA.

Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology*, *54*, 579–616.

Bolger, N., & Laurenceau, J.-P. (2013). *Intensive longitudinal methods: An introduction to diary and experience sampling research*. New York, NY: Guilford Press.

Bolger, N., Stadler, G., & Laurenceau, J.-P. (2012). Power analysis for intensive longitudinal studies. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 285–301). New York, NY: Guilford Press.

Bonnemeier, H., Wiegand, U. K. H., Brandes, A., Kluge, N., Katus, H. A., Richardt, G., & Potratz, J. (2003). Circadian profile of cardiac autonomic nervous modulation in healthy subjects: Differing effects of aging and gender on heart rate variability. *Journal of Cardiovascular Electrophysiology*, *14*, 791–799. DOI: 10.1046/j.1540-8167.2003.03078.x

Borkenau, P., & Liebler, A. (1995). Observable attributes as manifestations and cues of personality and intelligence. *Journal of Personality*, *63*, 1–25.

Borkenau, P., Mauer, N., Riemann, R., Spinath, F. M., & Angleitner, A. (2004). Thin slices of behavior as cues of personality and intelligence. *Journal of Personality and Social Psychology*, *86*, 599–614.

Boynton, M. H., & Richman, L. S. (2014). An online daily diary study of alcohol use using Amazon's Mechanical Turk. *Drug and Alcohol Review*. DOI: 10.1111/dar.12163

Bradburn, N. N., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design*. San Francisco: Jossey-Bass.

Brockner, J. (1979). Self-esteem, self-consciousness, and task performance: Replications, extensions, and possible explanations. *Journal of Personality and Social Psychology*, *37*, 447–461.

Brose, A., Lindenberger, U., & Schmiedek, F. (2013). Affective states contribute to trait reports of affective well-being. *Emotion*, *13*, 940–948. DOI: 10.1037/a0032401

Brosschot, J. F., & Thayer, J. F. (1998). Anger inhibition, cardiovascular recovery, and vagal function: A model of the link between hostility and cardiovascular disease. *Annals of Behavioral Medicine*, *20*, 326–332.

Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. Berkeley: University of California Press.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk : A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science*, *6*, 3–5. DOI: 10.1177/1745691610393980

Burg, M. M., & Pickering, T. G. (2011). The cardiovascular system. In R. J. Contrada (Ed.), *The handbook of stress science: biology, psychology, and health* (pp. 37–45). New York: Springer.

Bussmann, J. B. J., Ebner-Priemer, U. W., & Fahrenberg, J. (2009). Ambulatory activity monitoring: Progress in measurement of activity, posture, and specific motion patterns in daily life. *European Psychologist*, *14*, 142–152.

Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (Eds.), (2007). *Handbook of psychophysiology* (2nd ed.). Cambridge, England: Cambridge University Press.

Cervone, D. (2005). Personality architecture: Within-person structures and processes. *Annual Review of Psychology*, *56*, 423–452.

Chittaranjan, G., Blom, J., & Gatica-Perez, D. (2013). Mining large-scale smartphone data for personality studies. *Personal and Ubiquitous Computing*, *17*, 433–450.

Chui, H., Hoppmann, C. A., Gerstorf, D., Walker, R., & Luszcz, M. A. (2014). Social partners and momentary affect in the oldest-old: The presence of others benefits affect depending on who we are and who we are with. *Developmental Psychology*, *50*, 728–740. DOI: 10.1037/a0033896

Colvin, C. R. (1993). Judgeable people: Personality, behavior, and competing explanations. *Journal of Personality and Social Psychology*, *64*, 861–873.

Conner, T. S. (2014). *Experience sampling and ecological momentary assessment with mobile phones*. Retrieved from http://www.otago.ac.nz/psychology/otago047475.pdf

Conner, T., Barrett, L. F., Tugade, M. M., & Tenner, H. (2007). Idiographic personality: The theory and practice of experience sampling. In R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality research* (pp. 79–96). New York: Guildford Press.

Conner, T., & Lehman, B. J. (2012). Getting started: Launching a study in daily life. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 89–109). New York, NY: Guilford Press.

Conner, T., & Mehl, M. R. (in press). Ambulatory assessment—Methods for studying everyday life. In R. A. Scott, S. M. Kosslyn & M. Stephen (Eds.), *Emerging Trends in the Social and Behavioral Sciences*. Thousand Oaks, CA: SAGE Publications.

Conner, T. S., & Reid, K. A. (2012). Effects of intensive mobile happiness reporting in daily life. *Social Psychological and Personality Science*, *3*, 315–323. DOI: 10.1177/1948550611419677

Cooper, W. H., & Withey, M. J. (2009). The strong situation hypothesis. *Personality and Social Psychology Review*, *13*, 62–72. DOI: 10.1177/1088868308329378

Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion*, *12*, 307–330.

D'Arcey, J. T. (2013). *Assessing the validity of FaceReader using facial EMG* (Master's thesis, California State University). Retrieved from http://hdl.handle.net/10211.4/610

de Montjoye, Y.-A., Quoidbach, J., Robic, F., & Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. In *Social computing, behavioral-cultural modeling and prediction* (pp. 48–55). Berlin: Springer.

de Montjoye, Y.-A., Wang, S. S., Pentland, A., Anh, D. T. T., & Datta, A. (2012). On the trusted use of large-scale personal data. *IEEE Data Engineering Bulletin*, *35*, 5–8.

Denissen, J. J. A., & Penke, L. (2008). Motivational individual reaction norms underlying the Five-Factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, *42*, 1285–1302. DOI: 10.1016/j.jrp.2008.04.002

Diamond, L. M., & Otter-Henderson, K. D. (2007). Physiological measures. In R. W. Robins, C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 370–388). New York: Guilford Press.

Do, T., & Gatica-Perez, D. (2013). *The places of our lives: Visiting patterns and automatic labeling from longitudinal smartphone data. No. EPFL-ARTICLE-192391*.

Dufner, M., Arslan, R. C., Hagemeyer, B., Schönbrodt, F. D., & Denissen, J. J. A. (2014). Affective contingencies in the affiliative domain: Physiological assessment, associations with the affiliation motive, and prediction of behavior. *Journal of Personality and Social Psychology*.

Eagle, N., & Pentland, A. (2005). Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, *10*, 255–268. DOI: 10.1007/s00779-005-0046-3

Ebner-Priemer, U. W., & Kubiak, T. (2007). Psychological and psychophysiological ambulatory monitoring: A review of hardware and software solutions. *European Journal of Psychological Assessment*, *23*, 214–226.

Ekman, P., & Friesen, W. V. (1978) *Facial Action Coding System*. PaloAlto, CA: Consulting Psychologist Press.

Epstein, D. H., Tyburski, M., Craig, I. M., Phillips, K. A., Jobes, M. L., Vahabzadeh, M.,…Preston, K. L. (2014). Real-time tracking of neighborhood surroundings and mood in urban drug misusers: Application of a new method to study behavior in its geographical context. *Drug and Alcohol Dependence*, *134*, 22–29. DOI: 10.1016/j.drugalcdep.2013.09.007

Eyberg, S. M., Nelson, M. M., Duke, M., & Boggs, S. R. (2009). *Manual for the dyadic parent–child interaction coding system* (3rd ed.). Retrieved October 20, 2014 from http://pcit.phhp.ufl.edu/Measures.htm

Finkel, E. J., & Eastwick, P. W. (2008). Speed-dating. *Current Directions in Psychological Science*, *17*, 193–197.

Finkel, E. J., Eastwick, P. W., & Reis, H. T. (2015). Best research practices in psychology: Illustrating epistemological and pragmatic considerations with the case of relationship science. *Journal of Personality and Social Psychology*, *108*, 275–297. DOI: 10.1037/pspi0000007

Finkel, E. J., Eastwick, P. W., & Matthews, J. (2007). Speed-dating as an invaluable tool for studying romantic attraction: A methodological primer. *Personal Relationships*, *14*, 149–166.

Fleeson, W. (2001). Toward a structure-and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, *80*, 1011–1027. DOI: 10.1037/pspi0000007

Fleeson, W., & Jayawickreme, E. (in press). Whole trait theory. *Journal of Research in Personality*. DOI:10.1016/j.jrp.2014.10.006

Foley, P., & Kirschbaum, C. (2010). Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings. *Neuroscience & Biobehavioral Reviews*, *35*, 91–96.

Fraley, R. C. (2004). *How to conduct behavioral research over the Internet: A beginner's guide to HTML and CGI/Perl*. New York: Guilford.

Frauendorfer, D., Mast, M. S., Nguyen, L., & Gatica-Perez, D. (2014). Nonverbal social sensing in action: Unobtrusive recording and extracting of nonverbal behavior in social interactions illustrated with a research example. *Journal of Nonverbal Behavior*, *38*, 231–245.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review*, *102*, 652–670.

Funder, D. C. (1999). *Personality judgment: A realistic approach to person perception*. San Diego, CA: Academic Press.

Funder, D. C. (2006). Towards a resolution of the personality triad: Persons, situations, and behaviors. *Journal of Research in Personality*, *40*, 21–34. DOI: 10.1016/j.jrp.2005.08.003

Funder, D. C., Furr, R. M., & Colvin, C. R. (2000). The Riverside Behavioral Q-sort: A tool for the description of social behavior. *Journal of Personality*, *68*, 451–490.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology*, *64*, 479–490.

Furr, R. M., & Funder, D. C. (2009). Behavioral observation. In R. W. Robins, C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 273–291). New York: Guilford Press.

Gebauer, J. E., Leary, M. R., & Neberich, W. (2012). Big two personality and big three mate preferences: Similarity attracts, but country-level mate preferences crucially matter. *Personality and Social Psychology Bulletin*, *38*, 1579–1593. DOI: 10.1177/0146167212456300

Geisler, F. C. M., & Kubiak, T. (2009). Heart rate variability predicts self-control in goal pursuit. *European Journal of Personality*, *23*, 623–633.

Geukes, K., Hutteman, R., Küfner, A. C. P., Nestler, S., & Back, M. D. (2015). *Explaining the longitudinal interplay of personality and social relationships in the laboratory and in the field: The PILS and CONNECT study. Manuscript in preparation*.

Goodwin, M. S. (2012). Passive telemetric monitoring: Novel methods. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 251–266). New York, NY: Guilford Press.

Gonzaga, G. C., Campos, B., & Bradbury, T. (2007). Similarity, convergence, and relationship satisfaction in dating and married couples. *Journal of Personality and Social Psychology*, *93*, 34–48.

Gordis, E. B., & Margolin, G. (2001). The Family Coding System: Studying the relation between marital conflict and family interaction. In P. K. Kerig & K. M. Lindahl (Eds.), *Family observational coding systems: Resources for systemic research*. (pp. 111–125). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology*, *74*, 1337–1349.

Gosling, S. D., & Johnson, J. A. (2010). *Advanced methods for conducting online behavioral research*. Washington, DC: American Psychological Association.

Gunthert, K. C., & Wenze, S. J. (2012). Daily diary methods. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 144–159). New York, NY: Guilford Press.

Gustafson, D. H., McTavish, F. M., Chih, M. Y., Atwood, A. K., Johnson, R. A., Boyle, M. G.,…Shah, D. (2014). A smartphone application to support recovery from alcoholism a randomized clinical trial. *JAMA Psychiatry*, *71*, 566–572. DOI: 10.1001/jamapsychiatry.2013.4642

Hall, J., & Watson, W. H. (1970). The effects of a normative intervention on group decision-making performance. *Human Relations*, *23*, 299–317.

Hasler, B. P., Mehl, M. R., Bootzin, R. R., & Vazire, S. (2008). Preliminary evidence of diurnal rhythms in everyday behaviors associated with positive affect. *Journal of Research in Personality*, *42*, 1537–1546. DOI: 10.1016/j.jrp.2008.07.012

Hektner, J. M., Schmidt, J. A., & Czikszentmihalyi, M. (2007). *Experience sampling method: Measuring the quality of everyday life*. Thousand Oaks, CA: Sage.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral & Brain Sciences*, *33*, 61–83; discussion 83-135.

Hernandez, J., Li, Y., Rehg, J. M., & Picard, R. W.. (2014). *BioGlass: Physiological Parameter Estimation Using a Head-mounted Wearable Device*. Paper presented at the 4th International Conference on Wireless Mobile Communication and Healthcare, Athens, Greece.

Hewson, C., & Lason, D. (2008). Research design and tools for Internet research. In N. Fielding, R. M. Lee & G. Blank (Eds.), *Sage handbook of online research methods* (pp. 58–78). Thousand Oaks: Sage.

Hicks, J., Ramanathan, N., Falaki, H., Longstaff, B., Parameswaran, K., Monibi, M.,…Estrin, D. (2011). *ohmage: An open mobile system for activity and experience sampling*. Technical report, CENS Technical Reports, Submitted to ACM Transactions on Embedded Computing Systems.

Hirschmüller, S., Egloff, B., Schmukle, S. C., Nestler, S. & Back, M. D. (in press). Accurate judgments of neuroticism at zero acquaintance: A question of relevance. *Journal of Personality*. DOI: 10.1111/jopy.12097

Hofman, L. F. (2001). Human saliva as a diagnostic specimen. *The Journal of Nutrition*, *131*, 1621–1625.

Hofmann, W., & Patel, P. V. (2014). SurveySignal: A convenient solution for experience sampling research using participants' own smartphones. *Social Science Computer Review*. DOI: 10.1177/0894439314525117

Ickes, W. (1983). A basic paradigm for the study of unstructured dyadic interaction. *New Directions for Methodology of Social & Behavioral Science*, *15*, 5–21.

Ickes, W. (1993). Empathic accuracy. *Journal of Personality*, *61*, 587–610.

Ilmarinen, V.-J., Vainikainen, M.-P., Verkasalo, M., & Lönnqvist, J.-E. (2015). Why are extraverts more popular? Oral fluency mediates the effect of extraversion on popularity in middle childhood. *European Journal of Personality*, *29*, 138–151.

Intille, S. S. (2012). Emerging technology for studying daily life. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 267–282). New York, NY: Guilford Press.

Jacobs, R. S., Hyman, M. R., & McQuitty, S. (2001). Exchange-specific self-disclosure, social self-disclosure, and personal selling. *Journal of Marketing Theory and Practice*, *9*, 48–62.

Jennings, J. R., Kamarck, T., Stewart, C., Eddy, M., & Johnson, P. (1992). Alternate cardiovascular baseline assessment techniques: Vanilla or resting baseline. *Psychophysiology*, *29*, 742–750.

John, O. P. & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology*, *66*, 206–219.

Jokela, M., Bleidorn, W., Lamb, M. E., Gosling, S. D., & Rentfrow, P. J. (2015). Geographically varying associations between personality and life satisfaction in the London metropolitan area. *Proceedings of the National Academy of Sciences*, *112*, 725–730. DOI: 10.1073/pnas.1415800112

Kashy, D. A., & Kenny, D. A. (1990). Analysis of family research designs. *Communication Research*, *17*, 462–482.

Kenny, D. A. (1988). Interpersonal perception: A social relations analysis. *Journal of Social and Personal Relationships*, *5*, 247–261.

Kenny, D. A. (1994). *Interpersonal perception: A social relations analysis*. New York: Guilford Press.

Kenny, D. A. (2004). PERSON: A general model of interpersonal perception. *Personality & Social Psychology Review*, *8*, 265–280. DOI: 10.1207/s15327957pspr0803_3

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, *331*, 719–721. DOI: 10.1126/science.1197872

Kirschbaum, C., Pirke, K. M., & Hellhammer, D. H. (1993). The 'Trier Social Stress Test'–a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, *28*, 76–81.

Kogan, A., Oveis, C., Carr, E. W., Gruber, J., Mauss, I. B., Shallcross, A.,…Keltner, D. (2014). Vagal activity is quadratically related to prosocial traits, prosocial emotions, and observer perceptions of prosociality. *Journal of Personality and Social Psychology*, *107*, 1051–1063. DOI: 10.1037/a0037509

Kramer, A. D., Guillory, J. E., & Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*, *111*, 8788–8790. DOI: 10.1073/pnas.1320040111

Kranczioch, C., Zich, C., Schierholz, I., & Sterr, A. (2014). Mobile EEG and its potential to promote the theory and application of imagery-based motor rehabilitation. *International Journal of Psychophysiology*, *91*, 10–15.

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: A review. *Biological Psychology*, *84*, 394–421. DOI: 10.1016/j.biopsycho.2010.03.010

Kubiak, T., & Krog, K. (2012). Computerized sampling of experience and behavior. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life*. (pp. 124–143). New York, NY: Guilford Press.

Kudielka, B. M., Gierens, A., Hellhammer, D. H., Wüst, S., & Schlotz, W. (2012). Salivary cortisol in ambulatory assessment—some dos, some don'ts, and some open questions. *Psychosomatic Medicine*, *74*, 418–431.

Kurzius, E. & Borkenau, P. (2015). Antecedents and consequences of mimicry. *European Journal of Personality*, *29*, 107–124.

Lane, N. D., Lin, M., Mohammod, M., Yang, X., Lu, H., Cardone, G.,…Campbell, A. T. (2014). BeWell: Sensing sleep, physical activities and social interactions to promote well-being. *Journal of Mobile Networks and Applications*, *19*, 345–359.

Larson, G. M. & Sbarra, D. A. (2015). Participating in research on romantic breakups promotes emotional recovery via changes in self-concept clarity. *Social Psychologial and Personality Science*. DOI:10.1177/1948550614563085

Lathia, N., Pejovic, V., Rachuri, K. K., Mascolo, C., Musolesi, M., & Rentfrow, P. J. (2013). Smartphones for large-scale behavior change interventions. *Pervasive Computing, IEEE*, *12*, 66–73. DOI: 10.1109/MPRV.2013.56

Leiner, D. J. (2014). *SoSci Survey (Version 2.4.00-i) [Computer Software]*. Available from https://www.soscisurvey.de

Letzring, T. D. (2008). The good judge of personality: Characteristics, behaviors, and observer accuracy. *Journal of Research in Personality*, *42*, 914–932.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology*, *91*, 111–123. DOI: 10.1037/0022-3514.91.1.111

Littlewort, G., Whitehill, J., Tingfan, W., Fasel, I., Frank, M., Movellan, J., & Bartlett, M. (2011, 21-25 March 2011). *The Computer Expression Recognition Toolbox* (*CERT*). Paper presented at the Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on XXX.

Lovallo, W. R., & Thomas, T. L. (2007). Stress hormones in psychophysiological research: Emotional, behavioral, and cognitive implications. In J. T. Cacioppo, L. G. Tassinary, & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 342–367). Cambridge, England: Cambridge University Press.

Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T.,…Choudhury, T. (2012). *StressSense:*

DOI: 10.1002/per

Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 351-360). ACM.

Lu, H., Pan, W., Lane, N. D., Choudhury, T., & Campbell, A. T. (2009). *SoundSense: scalable sound sensing for people-centric applications on mobile phones*. Paper presented at the Proceedings of the 7th international conference on Mobile systems, applications, and services.

Luhmann, M., Orth, U., Specht, J., Kandler, C., & Lucas, R. E. (2014). Studying changes in life circumstances and personality: It's about time. *European Journal of Personality*, *28*, 256–266.

MacKerron, G., & Mourato, S. (2013). Happiness is greater in natural environments. *Global Environmental Change*, *23*, 992–1000.

Magnusson, M. S. (2000). Discovering hidden time patterns in behavior: T-patterns and their detection. *Behavior Research Methods*, *32*, 93–110.

Mairesse, F., Walker, M. A., Mehl, M. R., & Moore, R. K. (2007). Using linguistic cues for the automatic recognition of personality in conversation and text. *Journal of Artificial Intelligence Research*, *30*, 457–500.

Mehl, M. R., & Conner, T. S. (2012). *Handbook of research methods for studying daily life*. New York, NY: Guilford Press.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology*, *90*, 862–877. DOI: 10.1037/0022-3514.90.5.862

Mehl, M. R., & Holleran, S. E. (2007). An empirical analysis of the obtrusiveness of and participants' compliance with the electronically activated recorder (EAR). *European Journal of Psychological Assessment*, *23*, 248–257. DOI: 10.1027/1015-5759.23.4.248

Mehl, M. R., Pennebaker, J. W., Crow, D. M., Dabbs, J., & Price, J. H. (2001). The Electronically Activated Recorder (EAR): A device for sampling naturalistic daily activities and conversations. *Behavior Research Methods, Instruments, and Computers*, *33*, 517–523.

Mehl, M. R., Robbins, M. L., & Deters, F. G. (2012). Naturalistic observation of health-relevant social processes: The electronically activated recorder methodology in psychosomatics. *Psychosomatic Medicine*, *74*, 410–417. DOI: 10.1097/PSY.0b013e3182545470

Mehl, M. R., Vazire, S., Ramirez-Esparza, N., Slatcher, R. B., & Pennebaker, J. W. (2007). Are women really more talkative than men? *Science*, *317*, 82. DOI: 10.1126/science.1139940

Miller, G. (2012). The smartphone psychology manifesto. *Perspectives on Psychological Science*, *7*, 221–237. DOI: 10.1177/1745691612441215

Montoliu, R., Blom, J., & Gatica-Perez, D. (2013). Discovering places of interest in everyday life from smartphone data. *Multimedia Tools and Applications*, *62*, 179–207.

Morse, P. J., Sauerberger, K. S., Todd, E., & Funder, D. (2015). Relationships between personality, situational construal, and social outcomes. *European Journal of Personality*, *29*, 97–106.

Moskowitz, D. S., & Sadikaj, G. (2012). Event-contingent recording. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 160–175). New York, NY: Guilford Press.

Mund, M. & Neyer, F. J. (2014). Treating personality-relationship transactions with respect: Narrow facets, advanced models, and extended time frames. *Journal of Personality and Social Psychology*, *107*, 352–368. DOI:

Myrtek, M., Foerster, F., & Brügner, G. (2001). *Freiburger Monitoring System* (FMS). Frankfurt/M: Peter Lang.

Narayanan, S. S. G., Panayiotis, G. (2013). Behavioral signal processing: Deriving human behavioral informatics from speech and language. *Proceedings of the IEEE*, *101*, 1203–1233.

Nesselroade, J. R. (1991). The warp and woof of the developmental fabric. In R. Downs, L. Liben, & D. Palermo (Eds.), *Visions of development, the environment, and aesthetics: The legacy of Joachim F. Wohlwill*. (pp. 213–240). Hillsdale, NJ: Erlbaum.

Nestler, S., Grimm, K., & Schönbrodt, F. (2015). The social consequences and mechanisms of personality: How to analyze longitudinal data from dyadic, round-robin, and network designs. *European Journal of Personality*, *29*, 272–295.

Neyer, F. J., & Asendorpf, J. B. (2001). Personality-relationship transaction in young adulthood. *Journal of Personality and Social Psychology*, *81*, 1190–1204.

Neyer, F. J., Mund, M., Zimmermann, J., & Wrzus, C. (2013). Personality-relationship transactions revisited. *Journal of Personality*. DOI: 10.1111/jopy.12063

Onnela, J.-P., Waber, B. N., Pentland, A., Schnorf, S., & Lazer, D. (2014). Using sociometers to quantify social interaction patterns. *Scientific Reports*, *4*. DOI: 10.1038/srep05604

Ozer, D. J., & Benet-Martinez, V. (2006). Personality and the prediction of consequential outcomes. *Annual Review of Psychology*, *57*, 401–421.

Paolacci, G., & Chandler, J. (2014). Inside the turk: Understanding Mechanical Turk as a participant pool. *Current Directions in Psychological Science*, *23*, 184–188. DOI: 10.1177/0963721414531598

Paunonen, S. V. (2003). Big Five factors of personality and replicated predictions of behavior. *Journal of Personality and Social Psychology*, *84*, 411–422.

Qiu, L., Lin, H., Ramsay, J., & Yang, F. (2012). You are what you tweet: Personality expression and perception on Twitter. *Journal of Research in Personality*, *46*, 710–718. DOI: 10.1016/j.jrp.2012.08.008

Rachuri, K. K., Efstratiou, C., Leontiadis, I., Mascolo, C., & Rentfrow, P. J. (2013). *METIS: Exploring mobile phone sensing offloading for efficiently supporting social sensing applications*. Paper presented at the Pervasive Computing and Communications (PerCom), 2013 IEEE International Conference on.

Rachuri, K. K., & Mascolo, C. (2011), September. *Smartphone based systems for social psychological research: Challenges and design guidelines*. Paper presented at the 3rd ACM workshop on Wireless of the students, by the students, for the students.

Raento, M., Oulasvirta, A., & Eagle, N. (2009). Smartphones: An emerging tool for social scientists. *Sociological Methods & Research*, *37*, 426–454. DOI: 10.1177/0049124108330005

Rahman, T., Adams, A. T., Zhang, M., Cherry, E., Zhou, B., Peng, H., & Choudhury, T. (2014). *BodyBeat: A mobile system for sensing non-speech body sounds*. Paper presented at the Proceedings of the 12th annual international conference on Mobile systems, applications, and services.

Ram, N., Conroy, D. E., Pincus, A. L., Lorek, A., Rebar, A., Roche, M. J., & Gerstorf, D. (2014). Examining the interplay of processes across multiple time-scales: Illustration with the Intraindividual Study of Affect, Health, and Interpersonal Behavior (iSAHIB). *Research in Human Development*, *11*, 142–160. Doi: 10.1080/15427609.2014.906739

Ramanathan, N., Alquaddoomi, F., Falaki, H., George, D., Hsieh, C., Jenkins, J.,… Estrin, D. (2012), May. *ohmage: An open mobile system for activity and experience sampling*. In *Pervasive Computing Technologies for Healthcare* (*PervasiveHealth*), 2012 6th International Conference on (pp. 203-204). IEEE.

Ramírez-Esparza, N., Mehl, M. R., Álvarez-Bermúdez, J., & Pennebaker, J. W. (2009). Are Mexicans more or less sociable than Americans? Insights from a naturalistic observation study. *Journal of Research in Personality*, *43*, 1–7.

Rauers, A., Blanke, E., & Riediger, M. (2013). Everyday empathic accuracy in younger and older couples: Do you need to see your partner to know their feelings? *Psychological Science*, *24*, 2210–2217. DOI: 10.1177/0956797613490747

Rauthmann, J. F., Gallardo-Pujol, D., Guillaume, E. M., Todd, E., Nave, C. S., Sherman, R. A.,…Funder, D. C. (2014). The situational eight DIAMONDS: A taxonomy of major dimensions of situation characteristics. *Journal of Personality and Social Psychology*, *107*, 677–718. DOI: 10.1037/a0037250

DOI: 10.1002/per

Rauthmann, J., Sherman, R. A., & Funder, D. C. (2015). Principles of situation research: Towards a better understanding of psychological situations. *European Journal of Personality*, in press.

Reis, H. T. (2008). Reinvigorating the concept of situation in social psychology. *Personality and Social Psychology Review*, *12*, 311–329.

Reis, H. T. (2012). Why researchers should think "real-world": A conceptual rationale. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life*. (pp. 3–21). New York, NY: Guilford Press.

Reis, H. T., & Gable, S. L. (2000). Event sampling and other methods for studying everyday experience. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology*. (pp. 190–222). New York: Cambridge University Press.

Rentfrow, P. J., & Gosling, S. D. (2006). Message in a ballad: The role of music preferences in interpersonal perception. *Psychological Science*, *17*, 236–242.

Reitz, A. K., Zimmermann, J., Hutteman, R., Specht, J., & Neyer, F. J. (2014). How peers make a difference: The role of peer groups and peer relationships in personality development. *European Journal of Personality*, *28*, 279–288.

Riediger, M., Wrzus, C., Klipker, K., Müller, V., Schmiedek, F., & Wagner, G. G. (2014). Outside of the laboratory: Associations of working-memory performance with psychological and physiological arousal vary with age. *Psychology and Aging*, *29*, 103–114. DOI: 10.1037/a0035766

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socio-economic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313–345.

Robins, R. W. & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, *80*, 340–352.

Robins, R. W., Fraley, R. C., & Krueger, R. F. (Eds.). (2009). *Handbook of research methods in personality psychology*. New York, NY: Guilford Press.

Robinson, E. A., & Eyberg, S. M. (1981). The dyadic parent–child interaction coding system: Standardization and validation. *Journal of Consulting and Clinical Psychology*, *49*, 245–250. DOI: 10.1037/0022-006X.49.2.245

Sadikaj, G., Moskowitz, D. S., Zuroff, D. C. (2015). Felt security in daily interactions as a mediator of the effect of attachment on satisfaction with a romantic partner. *European Journal of Personality*, *29*, 187–200.

Sbarra, D. A. (2006). Predicting the onset of emotional recovery following nonmarital relationship dissolution: Survival analyses of sadness and anger. *Personality and Social Psychology Bulletin*, *32*, 298–312.

Schaffhuser, K., Allemand, M., & Martin, M. (2014). Personality traits and relationship satisfaction in intimate couples: Three perspectives on personality. *European Journal of Personality*, *28*, 120–133.

Schinka, J. A., & Velicer, W. F. (2012). *Handbook of Psychology: Research Methods*. (Vol. 2). I. B. Weiner (Ed.), Hoboken, NJ: Wiley.

Schönbrodt, F. D., & Perugini, M. (2013). At what sample size do correlations stabilize?. *Journal of Research in Personality*, *47*, 609–612.

Schlotz, W. (2012). Ambulatory psychoneuroendocrinology: Assessing salivary cortisol and other hormones in daily life. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 193–209). New York, NY: Guilford Press.

Schultheiss, O. C., & Stanton, S. J. (2009). Assessment of salivary hormones. In E. Harmon-Jones & J. S. Beer (Eds.), *Methods in Social Neuroscience* (pp. 17–44). New York: The Guilford Press.

Schwarz, N., & Oyserman, D. (2001). Asking questions about behavior: Cognition, communication, and questionnaire construction. *American Journal of Evaluation*, *22*, 127–160.

Scully, C., Lee, J., Meyer, J., Gorbach, A. M., Granquist-Fraser, D., Mendelson, Y., & Chon, K. H. (2012). Physiological parameter monitoring from optical recordings with a mobile phone. *Biomedical Engineering, IEEE Transactions on*, *59*, 303–306. DOI:10.1109/TBME.2011.2163157

Sherman, R. A., Nave, C. S., & Funder, D. C. (2013). Situational construal is related to personality and gender. *Journal of Research in Personality*, *47*, 1–14. DOI: 10.1016/j.jrp.2012.10.008

Shilton, K. (2009). Four billion little brothers?: Privacy, mobile phones, and ubiquitous data collection. *Communications of the ACM*, *52*, 4853.

Stappenbeck, C. A., & Fromme, K. (2014). The effects of alcohol, emotion regulation, and emotional arousal on the dating aggression intentions of men and women. *Psychology of Addictive Behaviors*, *28*, 10–19. DOI: 10.1037/a0032204

Stark, A., Shafran, I., & Kaye, J. (2014). Inferring social nature of conversations from words: Experiments on a corpus of everyday telephone conversations. *Computer Speech & Language*, *28*, 224–239. DOI: 10.1016/j.csl.2013.06.003

Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management, and meta-accuracy. *European Journal of Personality*, *28*, 73–94.

Sweeney, L. (1997). Weaving technology and policy together to maintain confidentiality. *Journal of Law, Medicine & Ethics*, *25*, 98–110.

Terzis, V., Moridis, C. N., & Economides, A. A. (2010). *Measuring instant emotions during a self-assessment test: the use of FaceReader*. In *Proceedings of the 7th International Conference on Methods and Techniques in Behavioral Research* (p. 18). ACM.

Trull, T. J., & Ebner-Priemer, U. (2014). The role of ambulatory assessment in psychological science. *Current Directions in Psychological Science*, *23*, 466–470.

Underwood, M. K., Rosen, L. H., More, D., Ehrenreich, S. E., & Gentsch, J. K. (2012). The blackberry project: Capturing the content of adolescents' text messaging. *Developmental Psychology*, *48*, 295–302.

Vater, A. & Schröder-Abé, M. (2015). Explaining the link between personality and relationship satisfaction: Emotion regulation and interpersonal behavior in conflict discussions. *European Journal of Personality*, *29*, 201–215.

Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, *98*, 281–300. DOI: 10.1037/a0017908

de Vries, S. I., Bakker, I., Hopman-Rock, M., Hirasing, R. A., & van Mechelen, W. (2006). Clinimetric review of motion sensors in children and adolescents. *Journal of Clinical Epidemiology*, *59*, 670–680.

Vogel, M., Monesson, A., Scott, L. S. (2012). Building biases in infancy: the influence of race on face and voice emotion matching. *Developmental Science*, *15*, 359–372.

Wagerman, S. A., & Funder, D. C. (2009). Situations. In P. J. Corr & G. Mathews (Eds.), *Cambridge handbook of personality* (pp. 27–42). Cambridge, England: Cambridge University Press.

Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S.,… Campbell, A. T. (2014). *StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones*. Paper presented at the Proceedings of the ACM Conference on Ubiquitous Computing.

Weiss, R. L., & Heyman, R. E. (1990). Observation of marital interaction. In F. D. Fincham & T. N. Bradbury (Eds.), *The psychology of marriage: Basic issues and applications*. (pp. 87–117). New York, NY US: Guilford Press.

Weiss, R. L., & Summers, K. J. (1983). Marital interaction coding system-III. In E. E. Filsinger (Ed.), *Marriage and family assessment: A source book for family therapy* (pp. 85–115). Beverly Hills, CA: Sage.

DOI: 10.1002/per

Westermann, R., Spies, K., Stahl, G., & Hesse, F. W. (1996). Relative effectiveness and validity of mood induction procedures: A meta-analysis. *European Journal of Social Psychology*, *26*, 557–580. DOI: 10.1002/(SICI)1099-0992

Wettstein, A., & Jakob, M. (2010). Assessing aggressive adolescents' environments from their perspective by using camera-glasses: An innovative new method. *Journal of Aggression, Conflict and Peace Research*, *2*, 23–32. DOI: 10.5042/jacpr.2010.0139

Wheeler, L., & Reis, H. T. (1991). Self-recording of everyday life events: Origins, types, and uses. *Journal of Personality*, *59*, 339–354.

Wilhelm, F. H., Grossman, P., & Müller, M. I. (2012). Bridging the gap between the laboratory and the real world: Integrative ambulatory psychophysiology. In M. R. Mehl & T. Connor (Eds.), *Handbook of research methods for studying daily life* (pp. 210–234). New York, NY: Guilford Press.

Wilhelm, P., & Perrez, M. (2004). How is my partner feeling in different daily-life settings? Accuracy of spouses' judgements about their partner's feelings at work and at home. *Social Indicators Research*, *67*, 183–246.

Wolf, P. S. A., & Jacobs, W. J. (2010). GPS technology and human psychological research: A methodological proposal. *Journal of Methods and Measures in the Social Sciences*, *1*, 1–5.

Wrzus, C., Müller, V., Lindenberger, U., Wagner, G. G., & Riediger, M. (2014). Affect dynamics across the lifespan: With age, heart rate reacts less strongly, but recovers more slowly from emotional strain. *Psychology and Aging*, *29*, 563–576. DOI: 10.1037/a0037451

Wrzus, C., Wagner, G. G., & Riediger, M. (2015). *Personality-situation-fit across the life span. Manuscript submitted for publication*.